Exploring the organisation of the L2 mental lexicon using sorting tasks.

Orita, Mitsuru

Exploring the Organisation of the L2 Mental Lexicon Using Sorting Tasks

Mitsuru Orita

A thesis submitted for the degree of Philosophiae Doctor

Swansea University

2009

DECLARATION

This work has not previously been accepted in substance for any degree and is not currently being submitted in candidature for any degree.

STATEMENT 1

This thesis is the result of my own investigation except where otherwise stated. Other sources are acknowledged by explicit references. A bibliography is appended.

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan and for the title and summary to be made available to outside organisations.

SIGNED ....

DATE ...5/01/2010...

Acknowledgements

I would first of all like to express my heartfelt gratitude to my supervisor, Professor Paul Meara, for his invaluable and painstaking guidance, profound wisdom and firm support throughout this thesis project. From him I have learned the richness of L2 vocabulary acquisition and the pleasure of working as a researcher in this field. His careful reading of the manuscript and feedback helped to substantially improve the final product.

I am deeply indebted to several friends and colleagues: Gail Braybrooks, Clarissa Wilks, Ellen Schur, Bill Perry, Josh Norman, Hidetoshi Kawatsu, Shiro Hasegawa and Kei Kobayashi. Gail, Clarissa, Ellen, and Bill read some of the earlier drafts of this thesis and Josh read the drafts in their entirety. Their insightful comments and helpful suggestions on many parts of this thesis made the final version more readable. Hidetoshi helped me with putting some of the data collected in the experiments into an analysable database. Shiro and Kei coached me on multivariate data analyses and helped me to analyse the data by applying cluster analysis and permutation tests. Without their support, this thesis could not have been completed.

I would also like to extend my deep thanks to Kazumi Aizawa, Fumio Seki, Fujiko Motoyama, Haruhiko Murao, Hitoshi Tokuda and Makoto Yoshii for their valuable advice and encouragement. My cordial thanks also go to students in the CALS PhD programme. I particularly thank Tadamitsu Kamimoto, Jim Ronald, Masamichi Mochizuki and Eugène Mollet for their comradeship and constructive comments on several aspects of this project. Expressions of gratitude are in order for my colleagues at Kumamoto University. In particular, I owe much to Richard Gilbert, Toshiyuki Kawamoto, Terry Laskowski, Alan Rosen, Timothy Stewart and Shunji Tsunoda. Their understanding and cooperation while I conducted my research were greatly appreciated. I must also thank the people who participated in the experiments of this project over the past seven years. Without their contributions, none of this work would have been possible.

I am heartily grateful to Masanori Yoshida, Hiroyuki Ito, Hirofumi Takeshita, Hiroya Okazaki, Yoshifusa Sato, Akihiko Higuchi, Ko-ichiro Hamasaki and Malcolm Benson. They have all introduced me to the value and delight of academic pursuits and often encouraged me while I was an undergraduate student as well as during the past 27 years while I have been teaching at various high schools and colleges. I would also like to sincerely acknowledge the unwavering friendship and warm encouragement that Shiro Ikeda, Masahiro Hori and Eiji Yamada have given me at various stages of my life and studies for over the past 30 years.

It is to my family that I owe more than I can probably ever repay. I thank Sayoko for her understanding, patience and support throughout my research. To my daughters Emina and Marina, I owe them much for the countless hours and days that I could have spent with them. I am also grateful to my late father-in-law, Morihide Sonoda, and mother-in-law, Chisa Sonoda, who have always been kind to my family. I cannot thank enough my father, Noriyuki Orita, and my late mother, Shizuko Orita, both of whom have always given me their unconditional love and understanding.

Abstract

This thesis is an attempt to probe into the organisation of the L2 mental lexicon through sorting tasks, which Haastrup and Henriksen (2000) pioneered. Particularly, it addresses whether the lexicons of native speakers of English and non-native speakers of English are different from each other quantitatively and qualitatively. After a replication of Haastrup and Henriksen, five experiments each using a different set of 50 high frequency English words taken from different parts of speech were conducted on 30 participants in each group (28 for the first experiment). The studies found that L1 and L2 differences were generally subtler than had been expected, such as for mean cluster number, cluster size, variability as well as the mean individual dendrogram and group dendrogram distances. However, cluster analysis showed that L2 lexical organisation was consistently different from L1 lexical organisation. Thus, it is highly plausible that the L2 mental lexicon has developed lexical networks that are on the surface similar to the L1 mental lexicon, when in fact the two lexicons have really developed different organisational structures from each other. Meanwhile, it was revealed that nouns can be predictors of L1 and L2 differences in all the tested variables. This result was attributed to the fact that the Japanese language has a significant number of loanword nouns that have originated from the English language. This suggests that these L2 lexical items, which are first learned as L1 lexical items as false friends, can be extremely difficult for Japanese L2 learners to re-learn and restructure into native-like L2 knowledge and organisation.

List of Figures

List of Tables

Contents

Chapter 1: Introduction

1.1 Introduction

Whether it is our first language or our second/foreign language, we access and use the "dictionary" in our brains in producing and understanding language. This dictionary in the brain is called the mental lexicon, which is defined as "a person's mental store of words, their meanings and associations" (Richards & Schmidt, 2002, p. 327). As this definition shows, the mental lexicon is characterised by its structure of associations (i.e., networks) in that words in the organisation are not in isolation but rather connected to each other with different strengths. The mental lexicon is essentially a gigantic, real structure. Aitchison (2003) states that the "relationship between a book dictionary and the human mental lexicon may be somewhat like the link between a tourist pamphlet advertising a seaside resort and the resort itself" (p. 14).

When we are learning a second language, our L2 mental lexicon is developing and is being organised in a different way from our L1 mental lexicon is (Aizawa, 2003a; Fitzpatrick, 2006; Habuchi, 2003; Henriksen, 1999; Kadota, 2003; Kroll &Tokowicz, 2001; Meara, 2001, 2002, 2007a; Singleton, 1999; Söderman, 1989, 1993). This thesis is an attempt at revealing the organisation of the L2 mental lexicon through sorting tasks, which Haastrup and Henriksen (2000) pioneered. Particularly, we will look into whether the L1 and L2 mental lexicons are different from each other in quantity and quality by analysing the clustering behaviours of L1 English speakers and L2 English speakers (advanced-level Japanese speakers of English). While analysing the results of whole experiments carried out in this project, this thesis also aims to reveal whether it is possible for L2 learners to attain native-like lexical knowledge and organisation. If the answer is "yes", it might imply that the L2 vocabulary learning and teaching environments currently in place are effective ones for language learners and teachers. Meanwhile, if the answer is "no", it would then be important to examine what aspects of L2 lexical knowledge and organisation are barriers to developing native-like knowledge and organisation. Either way, to the best of my knowledge, these are realms that have not been addressed in L2 vocabulary acquisition research thus far and that I believe will be worthwhile to address in this thesis.

1.2 Lexical knowledge and organisation

Dimensions of the mental lexicon are composed of vocabulary size (i.e., how many words one knows) and vocabulary organisation (how well words in the lexicon one has are structured). It is certainly a self-evident truth that vocabulary size is a crucial dimension for novice-level L2 learners to develop. With a limited L2 vocabulary, it is difficult for learners to understand and produce the target language efficiently and comfortably, even for basic and limited communicative purposes. It would be requisite for such learners to increase the size of their vocabulary at any cost. We cannot easily set a vocabulary size threshold that can be applicable to every L2 learner, but the high frequency words of about five thousand word families can be considered the minimum (Hirsh & Nation, 1992). When L2 learners have reached a certain threshold level, "vocabulary size per se seems to become less important" (Meara, 1996, p. 45). This suggests that by the time L2 learners have reached such a level, the interrelations among L2 words (i.e., lexical organisation) have also developed in some important ways. Meara (2004) has indicated that "vocabularies are not just collections of words, and that vocabularies are essentially interlocking networks" (p. 137). Considering that words in the brain are not in isolation, L2 learners who have achieved a certain threshold vocabulary level are expected to have developed a structured L2 lexical organisation in accordance with the development of their vocabulary size. Thus, how well L2 learners have structured their L2 lexical organisation could be a reliable way of determining whether they have developed a workable L2 vocabulary for efficient communication. Furthermore, by probing into the lexical organisation of advanced-level L2 speakers (and comparing it to the organisation of native speakers), we might be able to reveal whether such a desired restructuring and development of L2 lexical organisation has been attained.

Lexical knowledge is comprised of a number of facets. Nation (2001) provides a list of what is involved in knowing a word (Table 1.1), which allows us to grasp its facets comprehensively.

Table 1.1. What is involved in knowing a word

Form	spoken	R	What does the word sound like?
		P	How is the word pronounced?
	written	R	What does the word look like?
		P	How is the word written and spelled?
	word parts	R	What parts are recognisable in this word?
		P	What word parts are needed to express the meaning?
Meaning	form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	concepts and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	Associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	Grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	Collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	Constraints on use (register, frequency, etc.)	R	Where, when and how often would we expect to meet this word?
		P	Where, when, and how often can we use this word?

Note. R = receptive knowledge; P = productive knowledge. (Nation, 2001, p. 27)

Table 1.1 shows that lexical knowledge (i.e., knowing a word) is comprised of the three major facets: (a) word form, (b) meaning and (c) use. Furthermore, each facet is made up of three subcomponents: spoken, written and word parts for the facet of 'form', form and meaning, concepts and referents and associations for the facet of 'meaning' and grammatical functions, collocations and constraints on use for the facet of 'use'. In view of the relationship between lexical knowledge and lexical organisation and their development, the second facet of 'meaning' is particularly significant. Whether it is an L1 or L2, language learners usually go through all these three stages (i.e., form and meaning, concepts and referents and associations) in the 'meaning' facet in this order so as to develop a structured lexical organisation. Form-meaning mapping is a labelling task, concepts-referents mapping is a packaging task and associations between words are a network-building task. Nation (2001) states that these connections between each facet and task mirror what children go through in acquiring their L1 vocabulary (see also Aitchison, 1994, 2003). What matters is that L2 learners might take a different route in carrying out the network-building task and thus develop their L2 lexical organisation in a possibly different way from native-speaker organisation. For many L2 words that share the same concepts with their mother tongue, L2 learners might skip the first step of form-meaning mapping. For some L2 words, the packaging task might also be skipped since the L1 and L2 languages can be packaged under the same label. Thus, it is "clear that mature L2 learners do not experience the same mapping problems as young L1 learners who have to both develop concepts and learn to map words onto these concepts in the process of their cognitive development" (Henriksen, 1999, p. 308). It is predicted that L2 learners might develop a different lexical organisation from native speakers of the target language since their unique labelling and packaging tasks often affect their network-building process. As Henriksen points out, lexical development is characterised by its continuum nature, and during the process the labelling and packaging tasks that account for lexical item learning might have a substantial impact on the network-building of lexical organisation development and restructuring.

It should also be noted that both the lexical dimensions (size and organisation) and the tasks involved in organisational development (labelling, packaging and network-building tasks) continue to influence each other. For example, the bigger vocabulary size L2 learners have, the more interrelated their lexical networks will be. If the L2 vocabulary which learners are acquiring contains a large number of L1 cognates, then they may find it easy to carry out the task of labelling and packaging as well as network-building. Meanwhile, if the cognates are composed of 'false friends', the L2 learners' lexical organisation would be affected and the structure would be qualitatively different from its L1 counterpart. Thus it is speculated that the structure of L2 lexical organisation (which is the main concern of this thesis) may be different from that of L1 lexical organisation: not only the development and restructuring of L2 lexical knowledge and organisation but other factors (e.g., their L1 lexical knowledge and organisation) might also play a role.

Given the above discussion as a whole, it can safely be hypothesised that both lexical knowledge as item learning and lexical organisation as system development/restructuring in the L2 mental lexicon are not the same as those in the L1 mental lexicon. Furthermore, the steps that L2 lexical knowledge and organisation take in their formation seem to often be different from their L1 counterparts. Meara (1996) suggests that L2 lexical organisation is not as well structured as that of L1 organisation and that L2 speakers "find it less easy to produce associations than native speakers do, and are often unable to see connections between words that are obvious to native speakers" (p. 48). Meara argues that "a measure of this organisation might be a useful way of distinguishing between learners at different levels of proficiency" (p. 48). In the next section, we will briefly review the experimental studies that have addressed the issues of L2 lexical organisation.

1.3 Previous studies of L2 lexical organisation

In the past few decades, many important studies of L2 lexical organisation have been done using word association tests as a data elicitation technique (e.g., Kruse, Pankhurst & Sharwood-Smith, 1987; Nissen & Henriksen, 2006; Racine, 2008; Read, 1993; Schmitt, 2000;

Yokokawa, Yabuuchi, Kadota, Nakanishi & Noro, 2002). In the simplest form of the test (i.e., free word association test), participants are shown a series of single words and asked to tell or write the very first word they think of for each of the stimuli. The results are rich in information that discloses aspects of each participant's underlying lexical organisation. The stimulus-response patterns reflect the lexical organisation of the participants in that their relationships reveal the types of links in their mental lexicons. Thus word association tests are valid techniques in addressing L2 vocabulary size dimension (Meara & Fitzpatrick, 2000), distinguishing less proficient L2 learners from more proficient ones (Wolter, 2002) and illuminating the depth of lexical knowledge (Read, 2000).

A significant construct in WAT studies is response type. In syntagmatic associations, responses are words that could plausibly precede or follow the stimulus word in a sentence, as with the stimulus *mountain* eliciting the response *climb*, *walk* eliciting *slowly*, and *white* eliciting *snow*. In paradigmatic associations, the response could be substituted for the stimulus in a sentence. This includes co-ordinates (e.g., *green* → *yellow, son* → *daughter*), synonyms (e.g., *hard* → *difficult, old* → *ancient*), antonyms (e.g., *small* → *big, high* → *low*), super-ordinates (e.g., *lion* → *animal, tulip* → *flower*) and subordinates (e.g., *bird* → *pigeon, fruit*→ *apple*). The syntagmatic-paradigmatic shift is a widely confirmed phenomenon in WAT results between (a) young L1 children and adult native speakers and (b) novice L2 learners and advanced L2 learners/adult native speakers. Specifically, young L1 children tend to yield more syntagmatic responses than paradigmatic ones (Aitchison, 1994, 2003; Carter, 1987; Gass & Selinker, 1994, 2001; Hatch & Brown, 1995; McCarthy, 1990; Meara, 1980, 1983; Nelson, 1977). Similarly, novice L2 learners tend to produce more syntagmatic responses than paradigmatic ones compared to advanced L2 learners and native speakers. This can usually be interpreted as indicating that advanced L2 learners seem to have developed native-likeness in accordance with their development of the L2 vocabulary in size and depth, and that both advanced L2 learners and adult native speakers have constructed more meaning-based mental lexicons than young L1 children and novice L2 learners (Harley, 1995; Singleton, 1999; Söderman, 1989; Sökmen, 1993).

This syntagmatic-paradigmatic shift is usually explained in terms of the language proficiency of the subjects concerned, the way that their mental lexicons are organised in the brain, and their cognitive and intellectual development. Appel (1989), for example, claims that "it is generally assumed that the syntagmatic/paradigmatic shift is an expression of increasing linguistic and/or cognitive abilities, also because the shift was positively correlated with IQ" (p. 185). However, we cannot overlook the fact that even young L1 children will produce paradigmatic associations when stimulus words are drawn from certain word categories. For example, Folarin (1989) reports that young L1 children produced more paradigmatic responses than syntagmatic responses to 20 out of 24 stimulus words (concrete nouns such as *birds, fruits, insects*), with category name (i.e., nouns) being predominant. This study suggests that the kind of stimulus words selected can affect WAT results, bringing about peculiar response type distributions. See also Fitzpatrick (2006) for evidence showing the effect of stimulus words on WAT results.

Phonological associations are concerned with the stimulus-response pattern where the similarity in sound between the two words plays a role (e.g., *mother* → *other, need* → *needle, yellow* → *jello*). The phonological shift (a decline in frequency of clang associations between stimulus and response as learner proficiency increases) is another widely confirmed result in WAT experiments (Schmitt, 2000; Schmitt & Meara, 1997; Söderman, 1993). This shift is usually assumed to be due to the tendency of L1 children and less proficient L2 learners to seemingly depend on some sort of phonological link with the stimulus words (Carter, 1987). Phonological associations, however, could also be produced by wild guesses based on any clues that the sound or look of unfamiliar stimulus words might offer (Singleton, 1999).

Thus, as Fitzpatrick (2006) indicates, earlier studies of L2 lexical organisation using word association tests attempted to show that "L2 acquisition mirrors first language acquisition in that association preferences systematically shift from syntagmatic to paradigmatic." (p. 122). Earlier studies of L2 lexical organisation also attempted to confirm that there was a shift from phonology-oriented association patterns to semantics-oriented ones. However, as the above findings show, the syntagmatic-paradigmatic shift and the occurrence of fewer phonological associations are not always confirmed to be psycholinguistically real, and thus it is difficult to generalise these claims.

Another line of study addressing the syntagmatic-paradigmatic shift showed that not all words in the mental lexicon go through the same developmental shift according to the L2 proficiency of the learners (Söderman, 1993). This suggests that there might be a need to refine the methodology including the categorisation of WAT results. Orita (2000, 2002a, 2002b, 2003) reported that some stimulus words managed to evoke exceptionally frequent syntagmatic associations from the native speaker group and other stimulus words produced a very high frequency of paradigmatic associations from even the least proficient group. Nissen and Henriksen (2006) found that different word classes tend to produce different word association results. Using a set of words selected through a valid word selection process and categorisation of the results, Fitzpatrick (2006) revealed that a noticeable L1 and L2 difference is "the preponderance of defining synonym responses from native speakers as opposed to the high number of loose conceptual responses from non-native speakers" (p. 143). These studies seriously challenge the notion of the syntagmatic-paradigmatic shift in L2 lexical development.

It should also be noted that categorising word association results are not always easy and "there always seems to be a significant number of responses that cannot be classified with any degree of certainty, and this suggests that claims about a syntagmatic/paradigmatic difference in foreign-language learners need to be treated with caution" (Meara, 1980, p. 239).

It is true that word association tests as data elicitation techniques have been refined. Syntagmatic, paradigmatic and phonological associations are perhaps important aspects of L1 and L2 lexical organisation. The better structured lexicon is principally semantically structured. However, recent studies have also revealed that the widely-supported syntagmatic-paradigmatic shift is called into question and that syntagmatic aspects of the organisation also play a role in lexical development. Fitzpatrick's (2007) finding that "in direct contract to the findings of previous studies, native speakers generally produce more position-based responses (responses which in some way collocate with the cue word; similar to the conventional 'syntagmatic' classification) than non-native speakers" (p. 322) is evidence for this. It has been suggested that there are some universal differences in L1 and L2 lexical organisations, which may include differences in syntagmatic, paradigmatic and phonological links in the lexicons. Meanwhile, we cannot help feeling the limitations of free word association tests particularly in that their results need to be categorised by means of subjective judgment of the researchers concerned. By boosting the inter-rater reliability of categorisation in some way or another and fine-tuning the categorisation itself as Fitzpatrick (2006) did, the results can be valid and provide rich information in the lexical organisation. However, there may always be some difficulties in reliably categorising the WAT results. In the next section, we will look at another type of word association test where participants select a response to a stimulus among a limited set of words, which seems to solve this categorisation issue.

1.4 Unanswered issues

Another type of word association test is one which has participants select the strongest association pair from a given set of words. Thus, this is a restricted (forced) word association test, which is in sharp contract to the widely used free WAT where no limitation is imposed on the responses participants can produce. There are two types of restricted word association tests: one is the WAT where participants are asked to select any two words from a set of a handful of words (e.g., 40 trials in total) that they think are associated with each other (Wilks & Meara, 2002). The other is the WAT where participants select one word that best associates with the cue word among a pool of words (Meara & Schur, 2002). The merit of these restricted WATs is that the results can be analysed without needing any categorisation of the stimulus-response patterns researchers create and make judgment on. Wilks and Meara (2002), which addressed the issue of lexical density and did a numerical comparison of NS (native speaker) and NNS (non-native speaker) groups, revealed that the NS group had denser lexical organisation than the NNS counterpart. Meara and Schur (2002) addressed whether L1 participants were more aware of the lexical items belonging to distinct sets and their descriptive statistics-oriented analyses revealed that this was the case. (See Chapter 2 for a detailed review of Wilks & Meara (2002), Meara & Schur (2002) and other related studies, including their limitations.)

The point is that it is worthwhile for us to examine L2 lexical organisation by addressing its global structures further, instead of focusing on issues of syntagmatic, paradigmatic and phonological associations. As Wilks and Meara (2002) and Meara and Schur (2002) showed, L1 and L2 lexical organisations as a system are different from each other in crucial aspects. However, these studies leave some important questions unanswered. First, is L2 lexical organisation structured in the same way that L1 lexical organisation is structured? Studies using word association tests in general have shown that the lexical organisation of advanced L2 speakers is similar to native-like organisation in terms of the syntagmatic-paradigmatic shift. The question then becomes whether the L2 mental lexicon as a system is similar to the L1 mental lexicon when L2 speakers reach a highly proficient level. (Wilks and Meara predicted that L1 and L2 lexical organisations might be differently organised.) Second, are all the lexical networks in the L2 mental lexicon organised similarly regardless of word types (word classes)? Fitzpatrick (2006, 2007) and Nissen and Henriksen (2006) found that some word types are structured differently from others in the L1 mental lexicon. Is the same true with the L2 mental lexicon? Third, are L1 speakers more aware of lexical items belonging to distinct sets than their L2 counterparts? Meara and Schur showed this was the case using a restricted word association test. Can we, then, confirm this L1 and L2 difference in the degree of lexical awareness using another data elicitation task? This last question is related to my concern of whether another type of task besides a word association test still produces the same results that Meara and Schur's WAT-oriented study had revealed. Word association tests are primarily concerned with the participants' behaviour of stimulus-response patterns, and thus the test results reflect the underlying cognitive structures that are primarily composed of the relationships between two words. I contend that to explore L1 and L2 lexical organisation fully, another data elicitation technique should be introduced, where the task taps into the participants' behaviour beyond pairs of stimulus-response words. In this regard, sorting tasks look highly promising.

In their simplest form, sorting tasks involve having participants group a number of things (e.g., concepts, words, ideas, objects) into a smaller number of groups according to the similarity the participants think the words have with each other. Thus, sorting tasks "can be used to identify how concepts in a content area are organized in a learner's knowledge structure" (Jonassen, Beissner & Yacci, 1993, p. 45). The results reveal how concepts, words and other information in memory are related and what characteristics they share. When applied to lexical organisation research, words grouped (i.e., clustered) together can be labelled as lexical clusters in the structure. This is a distinct advantage over word association tests in that sorting tasks can tap into the links of more than two lexical items (i.e., clusters of lexical items) in the mental lexicon of the participants.

Sorting tasks have been adopted in many disciplines, such as medical studies, biology, anthropology, sociology, psychology and linguistics (see Coxon (1999) and Jonassen, Beissner & Yacci (1993) for a comprehensive review of sorting task-oriented studies). For example, addressing a rich diversity of semantic relations in natural language, Chaffin and Herrmann (1984) instructed 40 American college students to sort 31 cards, each of which had five example pairs of one of the 31 semantic relations on them. They found that the participants perceived five families of semantic relations: contrasts (e.g., *male-female*, *remember-forget*), class inclusion (e.g., *animal-horse*, *metal-copper*), similars (e.g., *car-auto*, *smart-intelligent*), case relationships (e.g., *artist-paint*, *dog-bark*) and part-wholes (e.g., *car-engine*, *tree-branch*). Chaffin and Herrmann also found that within each family, semantic relations were grouped in ways that were consistent with their defining properties.

However, sorting tasks have rarely been used in L2 lexical organisation studies. As discussed in section 1.3, almost all the significant findings and contributions to L2 lexical organisation research have been done by means of word association tests. Actually, Haastrup and Henriksen (1998, 2000) and Henriksen and Haastrup (1998) are the only studies that have addressed L2 organisation using sorting tasks. Considering that sorting tasks enable us to tap further into the cluster structures of lexical organisation than word association tests do, this thesis uses sorting tasks as a data elicitation technique.

1.5 Thesis outline

The following chapters document my attempt to explore the organisation of the L2 mental lexicon using sorting tasks. In Chapter 2, studies are reviewed that address L1 and L2 lexical organisation research, particularly working with the comparison of types of psycholinguistic data elicitation tasks and framing the research questions worth addressing in this thesis. In Chapter 3, a replication of Haastrup and Henriksen (2000) is reported and their sorting tasks are examined in light of whether they have validity as psycholinguistic data elicitation techniques. In Chapter 4, the results of the first full experiment using the sorting task I developed are reported. In Chapter 5, the results of the second experiment, which was done using a revised sorting task, are reported. Chapters 6 to 8 document the experiments that were done using further revised sorting tasks that were identical with each other except for the type of word tested in each task (verbs for Chapter 6, adjectives for Chapter 7 and nouns for Chapter 8). Through these five experiments, issues of L1 and L2 differences in lexical organisation, such as in cluster number, size and variability are examined. The examination includes whether the L1 and L2 lexical organisations are differently structured or not and which word class will be predictors of the L1 and L2 differences. In Chapter 9, reflection on the replication of Haastrup and Henriksen (2000) and all of the experiments carried out in the thesis are made. Issues that arose from the studies reported in the other chapters are also discussed. The discussion includes the question of the ultimate attainment of native-like lexical knowledge and organisation by L2 learners. Some closing comments are given in Chapter 10.

Chapter 2: Literature Review

2.1 Introduction

In Chapter 2, I will review the literature on L1 and L2 lexical organisation studies. As stated in Chapter 1, the purpose of this thesis is to examine cluster structures of L1 and L2 lexical organisations in the mental lexicons and how the two networks are different from each other. I argue for employing a new data elicitation technique for word association tests which are widely used in L2 lexical organisation studies. Therefore, in the first part of this chapter, I will review recent studies (those published since 2000) of L2 lexical organisation that have used word association tests and other tasks. They include a sorting task and similarity ratings.

In sections 2 to 5, I will review five papers which deal with lexical organisation in L2 speakers. The papers reviewed are Haastrup and Henriksen (1998), Henriksen and Haastrup (1998), Haastrup and Henriksen (2000), Wilks and Meara (2002) and Wilks, Meara and Wolter (2005). These papers provide the basic framework for the empirical work reported in later chapters. This section will also review two other studies, Meara and Schur (2002) and Sánchez (2004), which deal with lexical networks in L2 learners.

The review chapter will also look in detail at some work on L1 lexical networks (in sections 7 to 10). These papers include Miller (1969), Rapoport and Fillenbaum (1972), Preece (1976) and Routh (1994). These papers were selected because they introduce interesting methodological innovations in the study of lexical networks. These innovations have not yet been applied to L2 learners, but there is, in my view, much to be learned from them.

Each section will provide an objective summary of the paper under review, followed by a detailed critical evaluation of it. A number of common issues will emerge from these evaluations, and these issues will be discussed further in the discussion section.

2.2 Haastrup and Henriksen (1998), Henriksen and Haastrup (1998) and Haastrup and Henriksen (2000)

Haastrup and Henriksen (1998) and Henriksen and Haastrup (1998) were preliminary studies which are reported in more detail in Haastrup and Henriksen (2000). This section will therefore concentrate mainly on the last paper, while referring to the earlier studies as needed.

2.2.1 Summary

Haastrup and Henriksen (2000) investigated the developmental process of network building in

L2 learners' mental lexicons. The study is based on the theoretical model of L1 vocabulary acquisition developed by Aitchison (1994) while taking into consideration the formation of a lexical network in the mental lexicon. It also points to Meara (1996), who stresses the importance of lexical organisation for language performance, and Henriksen (1999), who proposes dimensions of L2 lexical knowledge and development. They argue that vocabulary acquisition is a matter of system learning rather than item learning while also supporting Ringbom (1983) and Henriksen (1996).

Haastrup and Henriksen (2000) attended to the lexical relationship between adjectives established by Miller and Fellbaum (1991): synonymy, antonymy and gradation. They addressed the issues of L2 learners' relational knowledge by a longitudinal research design. They proposed three hypotheses:

1. Being able to distinguish adjectives belonging to a basic emotion from adjectives of physical dimension is easier than placing adjective x into one of four lexical subsets of basic emotions (HAPPY, SCARED, SAD, ANGRY).
2. Being able to tell the difference in intensity between adjective x and adjective y is more difficult than knowing the basic emotion to which the two adjectives belong.
3. Finding the appropriate place in a lexical field for a particular adjective is a gradual and slow process and little development will be found across time.

(Haastrup & Henriksen, 2000, p. 227)

The participants were 34 12-year-old Danish school children. They were "near-beginners" (Haastrup & Henriksen, 2000, p. 228) who had been taught English for one and half years. Over a three-year period of the longitudinal study, 17 of them completed all the four tasks below. During the period, data was collected at one-year intervals: T1, T2 and T3. Over the same period, a group of British teenagers of the same age participated in the experiment. They provided a reference corpus to establish scoring keys to the tasks. In the Primary Sorting Task, participants placed 39 adjectives of emotion and physical dimension into one of four categories: HAPPY, AFRAID, WEIGHT and SIZE, or TEMPERATURE. In the Card-Sorting Task, they sorted 30 words into four subcategories of adjectives of emotion (SAD, HAPPY, ANGRY, AFRAID). After completion, they were asked to label each subcategory they had made. In the Gradation Task, participants graded a set of nearly synonymous adjectives selected from two lexical fields, HAPPY and AFRAID. In the paper, Haastrup and Henriksen did not include the results of the Gradation Task. This is because they "unintentionally made the Gradation Task easier than the Card-Sorting Task" (p. 229). In the Situation Task, participants selected near-synonyms to the word "that expressed the same basic emotion as the one intended" (p. 230) from a set of adjectives in 16 different situations.

To test Hypothesis 1, they compared the results of the Primary Sorting Task to those of the Card-Sorting Task at T2. The results revealed that, for five of the seven adjectives, participants performed better in the Primary Sorting Task. They stated that this tendency is more obvious for adjectives of AFRAID than those of HAPPY. Also, all the participants correctly sorted the word *glad*, which was attributed to the fact that the adjective is a high frequency word and is a true cognate between English and Danish. Another analysis using an implicational scale showed that if a participant could complete the Card-Sorting Task successfully, she[1] could accomplish the Primary Sorting Task as well. The researchers argued that, in order to successfully complete the Card-Sorting Task, "learners must have a well-developed network of paradigmatic relations in the form of synonymy" (p. 231).

To test Hypotheses 2 and 3, Haastrup and Henriksen did a case study. They examined the data obtained from two boys for the Card-Sorting Task and the Situation Task at T2 and T3. This was done because only these two participants revealed substantial lexical development over time. Both participants had improved performance in sorting adjectives of AFRAID and HAPPPY, but did not do so with SAD and ANGRY. It was argued that lexical input from crime and horror movies, thrillers and computer games might have played a role in the improved performance. The words *chuffed, furious* and *miserable* were the most difficult for the two boys. They are all low frequency words and have no formal similarity between English and Danish. The development of lexical knowledge was first observed in the correct answers in the Situation Task in T2 and then in the Card-Sorting Task in T3. Haastrup and Henriksen stated that network building is not a straightforward linear process but an extremely slow and gradual one.

The paper concluded that learners can develop their L2 vocabulary by focusing on lexical relationships and the analysis of words and that in general the findings support network building and depth of knowledge hypotheses in L2 lexical development.

2.2.2 Commentary

Haastrup and Henriksen (2000) was a pioneering study that probed into issues of L2 lexical organisation and development using sorting tasks. They opened up a new door to L2 lexical organisation studies by introducing sorting tasks. However, as Haastrup and Henriksen noticed, the results they obtained were mixed and rather weak to make firm conclusions and generalisations. Particularly, as predicted, the process of network building was very difficult to detect even through a longitudinal study. As the researchers had to work exclusively with only two boys out of the original 17 participants, it was not easy to find a distinct change in L2 lexical development. The importance of the research questions they addressed and the findings

[1] For generic reference third-person singular pronouns, *she* will be used in this thesis.

they made are well noted, but Haastrup and Henriksen (2000) has limitations in its research design. Among them, I'd like to discuss the problems in (a) their Card-Sorting Task and (b) the assumptions they made about semantic fields.

The Card-Sorting Task is the key task in which the network properties of the L2 mental lexicon should be represented. In the task, participants were asked to sort a pack of 30 cards (on which each one was printed one adjective of emotion) into groups. The results should have revealed how the words were related to each other in the L2 mental lexicon. Unfortunately, this was not the case. In the Card-Sorting Task, participants placed words into pre-determined sets of "correct basic emotions" (p. 229). They were SAD, HAPPY, ANGRY and AFRAID. However, if they did not know the meaning of the tested adjectives, participants could not complete the sorting task satisfactorily. The results reflected participants' lexical knowledge of synonyms and near-synonyms more than the degree of the development of network building. The results were less likely to reliably reflect underlying lexical structures.

Among the tested words, there were several lexical items that participants did not know the meanings of, even at later stages of the longitudinal study (e.g., *petrified, grumpy, chuffed*). They were low frequency words that many near-novice Danish participants did not master even at the final stage of the experiment. Unknown words among the tested words lessened the task reliability to probe into network-building issues. If participants had little idea of the word meanings, they could not have been expected to work out the relationship between them. For an experiment that aims to reveal facets of sorting behaviours and network building, the meanings of all the tested words need to be known to participants.

A sorting task that contains words unknown to participants embodies an aim different from the original research purpose. The results are more a test of knowledge of individual lexical items than a test of network building. In this regard, see Table 2.1 for a tabulation of scoring keys to the Card-Sorting Task established based on the answers of British informants on the task.

Table 2.1. Scoring keys based on British informants for Card-Sorting Task (Haastrup & Henriksen, 2000, p. 231)

AFRAID	ANGRY	HAPPY	SAD
alarmed	*annoyed*	*cheerful*	*depressed*
anxious	*cross*	*chuffed*	*disappointed*
distressed	*furious*	*excited*	*miserable*
frightened	*grumpy*	*glad*	*moody*
panic-stricken	*mad*	*high*	*sorrowful*
petrified	*moody*	*overjoyed*	*uneasy*
scared	*outraged*	*pleased*	*upset*
terrified		*thrilled*	
uneasy			

The table shows that some L1 informants sorted *uneasy* into AFRAID and others sorted it into SAD. This suggests that it is not easy for either L1 or L2 speakers of English to sort polysemous words into predetermined distinctive categories. Such a Card-Sorting Task could have been improved. One feasible solution to the problem would have been to give up scoring the L2 sorting task results against the scoring keys established by the L1 results. Instead, L1 and L2 sorting task results could have been simply compared to each other. With this revision, the issue of ambiguous scoring of the results, as in the case with *uneasy*, would have been resolved. The comparison would have revealed the extent to which L2 participants had developed these words into a network that was different from their L1 counterparts. Then the tasks would have tapped into the features of L2 network building rather than those of individual lexical knowledge.

Another limitation of the study is concerned with the assumption Haastrup and Henriksen (2000) made about network building. It might be true that L2 learners "build up semantic fields by adding terms to them as they elaborate their vocabularies and by creating links between words they already know and new L2 words" (pp. 221-222). However, this assumption is more related to item-learning of individual words than the organisational development of the whole lexicon. Lexical knowledge regarding semantic fields is certainly a crucial property of the network structure of the L2 mental lexicon, but whether the data obtained by their experiment offers relevant information about the latent cognitive system and development should be called into question. As indicated earlier, the results of their longitudinal study were affected significantly by the degree of the participants' knowledge of individual lexical items. While Haastrup and Henriksen's main goal was to study the process of network building, the data had only limited potential to answer their original research questions.

As the study reported, the development of network building in L2 mental lexicons can be a gradual and slow process. However, it is extremely difficult to prove such gradualness and slowness. The study found evidence of a change in only two participants. The results are too limited to be evidence for any broad claim in a study of 17 participants.

Haastrup and Henriksen (2000) is part of a comprehensive research project the researchers are conducting (Haastrup & Henriksen, 1998; Henriksen & Haastrup, 1998). It is expected that the research project as a whole will bring about firm evidence to support the claims of Haastrup and Henriksen (2000).

In conclusion, it should be stressed that the sorting tasks that Haastrup and Henriksen devised are not without merit. A sorting task has the validity to tap directly into the underling cluster structures of the mental lexicons. Such a task can be improved and employed in research on L2 lexical organisation.

2.3 Wilks and Meara (2002) and Wilks, Meara and Wolter (2005)

Wilks and Meara (2002) explored issues of L1 and L2 lexical network density. They applied mathematical principles of Graph Theory to word association test data, and made use of computer simulations to predict WAT experiment results. Wilks, Meara and Wolter (2005) is a follow-up study of Wilks and Meara (2002). I will review Wilks and Meara (2002), make a brief summary of the second study, and then comment on both studies.

2.3.1 Summary

Wilks and Meara (2002) focused on investigating the use of the network (cobweb) metaphor (Aitchison, 1987; Bogaards, 1994) in L2 lexical organisation studies. Wilks and Meara predicted that L1 lexicons are denser than L2 lexicons. For the examination, they used Graph Theory modelling. Wilks and Meara applied two properties of graphs to their study. One is that the paths between points become longer as the number of points in the graph increases. The other is that the paths become shorter in proportion to the increase of the number of connections each point has.

Wilks and Meara (2002) made a questionnaire in which L1 and L2 speakers of French answer perceived association pairs for sets of randomly selected words. They predicted that "given a set of random words, and asked to find any associations among them, L1 speakers would have a higher 'hit rate' than we would expect from L2 speakers" (Wilks & Meara, 2002, p. 310). Before testing it, they programmed and ran a computer (network) simulator that aimed at specifying a range of lexical network models by controlling two parameters. The parameters were the size of the lexicon and the number of connections (i.e., density) that each word in the lexicon has with other words. The size of the lexicon was established at 1,000 words, and the number of links between words was changed from 4 to 10, simulating vocabulary networks of increasing density. The network simulator used 40 items each consisting of a set of 5 randomly selected words with a group of 30 'pseudo-participants'. Each group of parameters was run 15 times on the simulator.

In the main study, 30 native speakers of French and 30 learners of French (L1 = English) at the college level participated. They completed a 40-item questionnaire that had the identical structure to the simulation experiment. Each item was made up of a set of five French words randomly selected from the first 1,000 most frequent words in the Français Fondamental list (Gougenheim, Michéa, Rivenc & Sauvageot, 1956). Participants were asked to circle any two words in each set that they thought were associated with each other. Where they perceived no connections between any of the words in the set, they wrote nothing. When they thought more than two of the words in the set to be linked, they circled the two with the strongest connection.

The results revealed that native speakers perceived significantly more associations than did L2 speakers. The average hit rate of native speakers was 30.90 and that of L2 speakers was 19.00 ($t = 6.47, p < .001$). Wilks and Meara (2002) interpreted this as showing that L1 mental lexicons were denser, having more linkage between items than did L2 mental lexicons. The mean hit rate turned out to be far higher for both L1 and L2 speakers than the simulator had predicted. Wilks and Meara attributed this discrepancy to two factors. First, their simulations underestimated the average number of links between items in both the L1 and L2 lexicons. Wilks and Meara re-ran the simulator to calculate the approximate hit rates they got by the real data. The results revealed that the link number parameter should have been raised to 36 links per word. This enabled them to gain mean hit rates of 19-20 associations per questionnaire. To gain mean hit rates of 30, the parameter had to be raised to more than 45 links per word. This was distinctly different from the findings of earlier word association-based L2 vocabulary studies that reported a lower density of L2 networks than did Wilks and Meara (2002). They argued that their present method of association recognition task had more sensitivity and led both L1 and L2 speakers to tap into a denser network of associations.

Wilks and Meara (2002) claimed that simple comparisons of average numbers of connections between L1 and L2 lexical networks may be more problematic and more misleading than generally assumed. They predicted that "two networks with the same density could in fact be quite differently arranged in terms of how the connections between points are disposed and how many isolated points are present in the network" (Wilks and Meara, 2002, p. 319).

Wilks, Meara and Wolter (2005) re-examined some of the assumptions of Wilks and Meara (2002). Then they reinterpreted the data to explain L1 and L2 word association behaviours more satisfactorily. Instead of defining associations as only those between words having strong links as proposed in the previous modelling, they adopted a looser assumption where associations between words frequently rely on a common link shared with a third word. They reprogrammed their simulator to "register a hit if any one of the words associated with word X also occurred in the list of associates for word Y" (Wilks et al., 2005, p. 364).

The revised simulator program worked well, and relatively high hit rates were obtained with relatively small numbers of links between words in the lexicon. That is, 11 links per word was sufficient to account for the L1 data and 7 links per word was sufficient to account for the L2 data. It was found that a difference between L1 and L2 does exist, but that the difference appeared to be much smaller. The researchers claimed that previous studies on L2 word association behaviours, including Wilks and Meara (2002), "may have been very naïve in assuming that word association behaviour was a direct reflection of the immediate connections between words" (Wilks et al., 2005, p. 371).

2.3.2 Commentary

Wilks and Meara (2002) and Wilks, Meara and Wolter (2005) are solid L2 lexical organisation studies in many respects. They both made a formal graph theoretical approach to word association data and employed computer simulations to predict L1 and L2 word association behaviours precisely. The researchers examined the data collected in real experiments against the simulation data. They further re-programmed and re-ran the original simulator to capture the L1 and L2 word association behaviours in a more reliable way. The two studies revealed that L1 and L2 mental lexicons were actually different from each other in regard to their density. I evaluated both of these studies very highly. Wilks and Meara (2002) and Wilks, Meara and Wolter (2005) are examples of in-depth and robust studies of L1 and L2 mental lexicons with a sophisticated use of computer simulations, but what strikes me most about the studies is their prediction that two lexical networks with the same level of density could be differently arranged. This is an area of lexical organisation research worthy of further investigation.

2.4 Meara and Schur (2002)

2.4.1 Summary

Meara and Schur (2002) considered randomly generated computer data simulating the word association behaviours of human participants. The results were compared with those produced by L1 and L2 participants. Thus, their focus was on whether "computer generated random networks served as a basis of comparison for the word association networks of native speakers and non-native speakers" (p. 169).

The study had two human participant groups; 32 adult bilinguals with L1 English and L2 Hebrew and 32 11th grade high school pupils whose L1 was Hebrew. The task was composed of 50 high frequency verbs that were put alphabetically in 'the Verb Box' and a randomly ordered list of the same 50 verbs in 'the Verb List'. In the experiment, participants were directed to select and write down another word from the verb box that they thought was best associated with it. They used a verb from the box as many times as they felt was needed. A computer programme made associations between the 50 verbs in the list, but did the task by selecting associations randomly.

The results were put into a group data (co-occurrence) matrix for each of the L1, L2 and random groups. Using the matrix data, Meara and Schur calculated the number of other words a stimulus word elicited as a response. Then each of the different responses was counted only once. For L1 speakers, each stimulus word produced an average of 8.5 different responses. For L2 speakers, it was 12.4 and for the random data, it was 23.8. An ANOVA revealed that these differences were statistically significant, especially between the random data and the two human data sets ($F\,(2, 49) = 662, p < .001$). A t-test disclosed that the difference between the L1 and L2 data was also significant ($t = 12.95, p < .001$). The L2 results lay somewhere between the L1 and random association results.

Regarding idiosyncrasy of association, the random group produced the highest idiosyncratic responses (responses generated by only one participant), followed by the L2 group and then the L1 group. With regard to the common responses (responses produced by more than a single participant) yielded by two to nine participants, the same order of the idiosyncratic responses was confirmed. Regarding the common responses produced by 10 or more participants, an inverse order was confirmed. Meara and Schur argued that the L1 group generated "a substantial number of common responses — for almost all of the stimulus words, more than 10 participants made the same association" (p. 172), and for the L2 group, "only a small number of common responses were produced by more than 10 participants" (p. 172). These results were consistent with those of previous studies in that L2 responses were more varied than L1 responses.

Meara and Schur (2002) argued that the results were "something of a problem" (p. 173). This was because L2 association networks are "denser" than L1 association networks. That is, the number of words that each word was associated with in the L2 results was larger than that in the L1 results. Moreover, the random group was the densest because almost all of the words were interconnected. Regarding this problem, they argued that the analyses based on the group data concealed the contribution each participant made. Therefore, individual networks were constructed and the properties of each network were examined. It was found that L1 networks were largely disconnected while L2 networks were more interconnected. Random networks usually consisted of one very large component. The random network data was very homogeneous, while both of the human data sets were fairly varied in terms of *SD* values. The results revealed that L2 participants produced more components of 16 or more nodes than L1 participants did. The random group predominantly produced a component having 16 or more nodes. The results were statistically significant ($\chi^2 = 62.9, p < .001$). Meara and Schur argued again that the notable differences were not between the L1 and L2 networks, but rather between the networks produced by the human groups and those produced by the random group.

In sum, the L1 lexical networks had many small non-connected components, and the random networks had only a few large interconnected components. The L2 participants fell between the two opposite poles. Meara and Schur (2002) argued that "non-native speakers, for the most part, do not yet perceive vocabulary as belonging to smaller, constrained and strongly connected sets and they tend to associate words in more diverse and less predictable ways than native speakers do. The native speaker networks seem to reflect the fact that L1 speakers have a greater awareness of the semantic relations between words, and how words may fall into distinct sets" (pp. 179-180).

Meara and Schur concluded that the computer-generated random networks they adopted made it possible to tap into the features of L2 lexical networks more easily. L2 lexical networks can be characterised as being distinctively different from typical computer-generated random networks. They argue that this approach to lexical organisation would not have been possible by conducting traditional experiments using L1 networks as a baseline.

2.4.2 Commentary

Meara and Schur (2002) showed that computer-generated random data plays a reliable role as a baseline measure in L1 and L2 lexical organisation studies. They also provided firm evidence that L2 networks are more varied and have fewer, larger connected components than L1 networks. Although the study contained these positive points, Meara and Schur (2002) still has some weaknesses. Among them are two problems: (a) the restricted word association test they developed and (b) their definition of lexical density.

First, the restricted word association test Meara and Schur developed has a limitation. It is true that a word association test of this sort has an advantage. Researchers do not get "wasted data" in that all the stimulus-response pairs can be mapped out. This has rarely been achieved in a traditional WAT experiment asking participants to produce a response that comes to mind. In free WAT experiments, researchers often encounter many stimulus-response samples that are idiosyncratic and difficult to categorise. In contrast, in Meara and Schur's (2002) approach, all pairs of stimulus-response could be mapped out. The results were easily made into a co-occurrence matrix, and the matrix made statistical analyses feasible. Actually, Meara and Schur confirmed important findings and evidence for previous L2 lexical organisation studies as summarised above. However, the restricted WAT required participants to select only one word and participants could use a verb as many times as they wanted. Accordingly, the variation of stimulus-response pairs was quite limited in number, particularly in the case of L1 participants. This constrained participants from revealing what their underlying lexical networks were really like. This restricted WAT has weak sensitivity in capturing the potential ways that a stimulus can be linked with other words. This limitation is identified in the figures that represent the first 12 words of a 50-word matrix of the L1, L2 and random groups (Meara & Schur, 2002, pp. 174-175). See Table 2.2 for a summary of the results.

Table 2.2. Cell types of stimulus-response pairs of the first 12 words of a 50-word matrix (adopted from Meara & Schur, 2002)

Cell type	L1 (*n* = 32)		L2 (*n* = 32)		Random (*n* = 32)	
	Number	%	Number	%	Number	%
Zero response	122	85%	119	83%	72	50%
One response	12	8%	15	10%	64	44%
2-5 responses	6	4%	7	5%	8	6%
6-9 responses	2	1%	1	1%	0	0%
10-19 responses	1	1%	2	1%	0	0%
20 or more responses	1	1%	0	0%	0	0%

Table 2.2 reveals that all three groups overwhelmingly produced zero or one response. In particular, both the L1 and L2 groups had more than 80% with zero responses, and 8 to 10% with only one response. These results contradict Meara and Schur's (2002) interpretation that the L1 English group generated "a substantial number of common responses — for almost all the stimulus words, more than 10 participants made the same association . . . and only a small number of common responses were produced by more than 10 [L2 English] subjects" (p. 172). This interpretation does not fully take into consideration the actual results represented in the group matrices. Perhaps the fundamental problem that brought about the incorrect interpretation lies in the restricted WAT itself which asked participants to produce a single association among 50 verbs. Schur (2007) proposed modifications of the restricted WAT used in Meara and Schur (2002). Modifications included increasing the number of stimulus words

and asking participants to identify two or three responses in completing the WAT. Given the results in Table 2.2, the revised WAT seems to be promising. With this revision, co-occurrence matrices which summarised group WAT results would be expected to be more filled-in while the number of blank cells would be expected to decrease.

Second, the definition of lexical density Meara and Schur (2002) adopted was not appropriately valid to be applied to the restricted WAT data. They regarded 'density' as a "reflection of the number of words each word is associated with, so that if the stimulus words generate lots of different associations, then the network for the group will be denser than if each word generates only one or two different associations" (p. 173). By applying this definition of density to the WAT data, the L2 results turned out to be denser than the L1 results, and the random group was the densest. Since the analysis was done on a group matrix, the L2 data had more filled-in cells than the L1 data and the random data had the most filled-in cells. This method of analysis is highly problematic. As stated earlier, the results had to do with group data analysis and masked what individual participant results were actually like. Moreover, the group data analysis concealed the strength of the associations and the number of participants who made a stimulus-response for each matrix cell. This is essential information in investigating L1 and L2 lexical organisations and their structural differences.

Put in another way, a large portion of the L1 associations were bi-directional. The L1 participants produced associated pairs using a more limited group of words than the L2 and random groups (Table 2.3). The L1 group produced less varied association results than the other groups. This affected the results when they did analyses with group matrices. The L1 matrix cells turned out to be less filled-in than the L2 and random data. The paper concluded that L1 lexical networks were not denser than L2 and random networks. Given the entire discussion, this is not the right way to do the analysis and is in conflict with the way individual L1 and L2 lexical networks are actually organised. As confirmed in the previous review of Wilks and Meara (2002) and Wilks, Meara and Wolter (2005), L1 lexical networks are denser than their L2 counterparts. The group data analysis Meara and Schur adopted is not appropriate for examining individual participant WAT behaviour and lexical organisation.

Table 2.3. Link types of L1, L2 and random networks (adopted from Meara & Schur, 2002)

	L1 (n = 32)		L2 (n = 32)		Random (n = 32)	
	Number	%	Number	%	Number	%
Single-directional link	17	50%	43	96%	47	100%
Bi-directional link	17	50%	2	4%	0	0%

With regard to the issues of random networks and modelling, attention should be drawn to a recent prediction: Meara's (2007b) prediction that random structures are not a good model for lexical networks and that human lexicons may exhibit the properties of a 'small world'. In small world networks, most nodes in the networks are connected to a small number of closely related nodes, and only a few connections go from one of these clusters to another. Meara noted that simulations by small world models generated data that look different from the data produced by random network models. Schur (2007) attempted further analysis of the present WAT results (including the results of two other groups). In the analysis, she paid closer attention to the small world properties and reported some promising results. Unfortunately, however, the results were rather mixed regarding differences of lexical organisation between participant groups. Particularly, one L2 learner group (Chinese) had WAT results that seemed to be affected by their vocabulary learning strategies. This suggests that word association behaviours are often affected by the culture and learning style participants have.

In spite of a few problems identified above, Meara and Schur (2002) is an insightful, scientific approach to L2 lexical organisation. Their random network model has some limitations as a baseline in lexical organisation studies. However, the restricted WAT asking participants to select the strongest association to a stimulus word succeeded in producing analysable data without any "wasted data." This is an outstanding merit of research of this sort. Having participants complete a psycholinguistic task with a limited set of words is a reliable way to investigate the lexical networks of real participants.

2.5 Sánchez (2004)

2.5.1 Summary

Sánchez (2004) claimed that L2 lexical structure would develop into that of a native speaker after instruction using L1 lexical organisation as a model. Sánchez worked with a set of English verbs taken from the semantic field *shine*. She addressed two research questions: (a) Are L2 learners' semantic maps more congruent with those of native speakers after instruction? and (b) Are L2 learners who receive experimental instruction more able to use target words appropriately than those who do not?

There were two participant groups in the experiment, the control and experimental groups, and each had 30 participants. They were all first year students of English Philology at a university in Spain. Thirty native speakers of English participated in the experiment as "experts" in English. The L1 data was submitted to the Pathfinder network algorithm to get a network representation of the underlying structure. The visualised representation was used as a model of "an expert mean net" (Figure 2.1) in the instruction given to the experimental group.

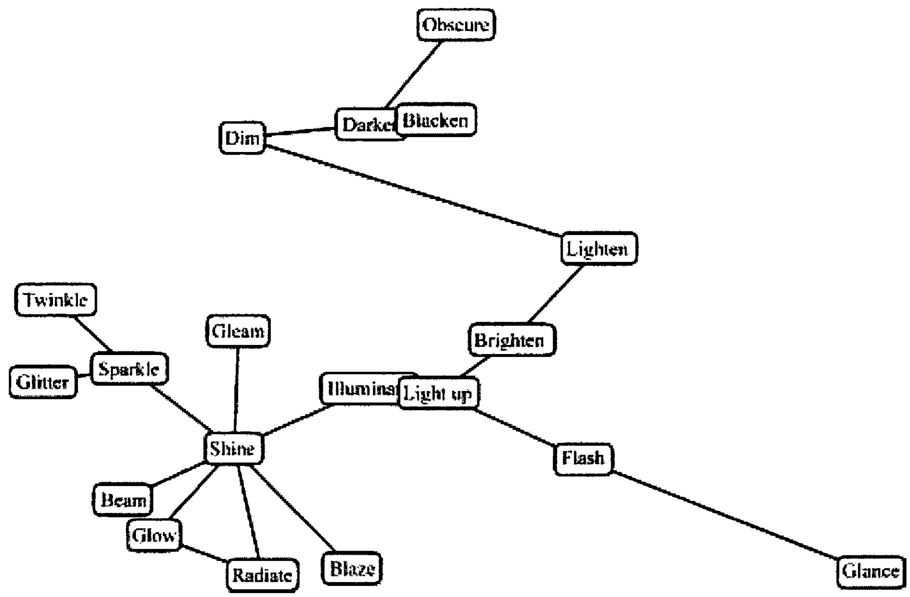

Figure 2.1. Expert mean net. The distance between the nodes indicates distance between the concepts. (Sánchez, 2004, p. 93)

Two different sessions of treatment instruction were given to the experimental group. The first session (10 minutes) was spent to "quickly familiarize the students with this vocabulary [of 19 *light* verbs]" (Sánchez, 2004, p. 93). The instruction focused on elaborating differences between the *light* verbs in intensity (e.g., *radiate* as more intense than *shine*) and situations of use (e.g., *twinkle* for stars and *glitter* for gold). In the second session (25 minutes), each student was given a copy of an "expert mean net". Using the handout, the teacher explained links that came out from each node (word), and discussed the proximities between words where a smaller distance meant a greater relatedness. Also, she explained the *light* verbs in English, presented them in context, and sometimes made use of example sentences in Spanish. Prior to the treatment instruction above, as pre-tests, similarity ratings and three kinds of lexical knowledge tests were given to both L2 groups. While doing the similarity ratings, participants rated relationships between each random pair of 19 *light* verbs that appeared on the computer screen. Test-takers pressed a key ranging from 1 to 9, with higher numbers representing greater relatedness. The results were subjected to the Pathfinder algorithm to construct a network representation. Lexical knowledge tests consisted of a multiple choice test and two fill-in-the-blank tests, using the semantic field *shine*. In the post-test two months later, the same tests were given to the L2 participants. The results showed that from *shine*, seven direct links with other words came in the expert network, as well as in the experimental group in the post-test. The expert net had only five links in common with the experimental group's pre-test net, but 15 common links with the post-test net. The control group showed no statistically significant difference between the results of pre- and post-lexical knowledge tests and those for similarity ratings. On the contrary, the lexical knowledge test results of the experimental group revealed that the gains in the post-test turned out to be statistically significant. Also, in the post-test, the experimental group significantly outweighed the control group in the results of their similarity ratings and lexical knowledge tests.

Given these results together, Sánchez (2004) concluded that instruction using an expert net was effective. She claimed that her treatment instruction had led L2 learners to get closer to experts in lexical organisation and to yield gains in the L2 lexical knowledge tests given.

2.5.2 Commentary

At first reading, the L1 "expert net" based instruction Sánchez (2004) contrived seemed promising in helping to develop native-like lexical structure. The results she reported supported her claims. However, a closer look at the paper revealed that it contained two serious limitations. They were that: (a) the role of a control group versus an experimental group, and (b) using an L1 "expert net" as a model in L2 lexical instruction. In addition, the merits and demerits of Pathfinder Analysis as a data analysis and representation technique will be discussed in this section.

First, Sánchez (2004) did not postulate the role of the control group in the experiment adequately enough to argue for the effect of the treatment instruction given to the experimental group. To claim that a treatment is effective, a study needs to report what the control group did during the period of treatment as compared to the experiment group. As Mackey and Gass (2005) states, to report what the experimental and control groups receive is "to ensure that it was the treatment, not the mere fact of doing something that led to any change" (p. 148). Sánchez failed to do so, thus her claim for the effectiveness of the treatment is decidedly weakened. Moreover, Sánchez did not mention whether her research conducted an immediate post-test to determine the on-the-spot effect of the treatment. Without being given immediate post-test results, it is difficult for the reader to decide whether the good results of a delayed post-test actually reflect a long-lasting effect of the treatment. The delayed post-test results might have been boosted by the effect of uncontrolled factors. It is plausible that participants in the experimental group were more conscious of *light* words after instruction. Thus, the results might have been produced by their awareness and retention of *light* words through subsequent exposure to English on TV programs and the Internet after instruction. It was also possible that participants remembered a limited set of the *light* words and the relationships they had learned in the instruction. Failing to report whether the researcher conducted an immediate post-test further degraded the reliability of the delayed post-test.

Second, the semantic mapping Sánchez (2004) contrived is severely lacking in validity. Semantic mapping, as widely accepted, involves the teacher and students working together to brainstorm. They jointly construct a visual link between lexical items and ideas on paper or the blackboard. An example from Nation (2001) is shown in Figure 2.2.

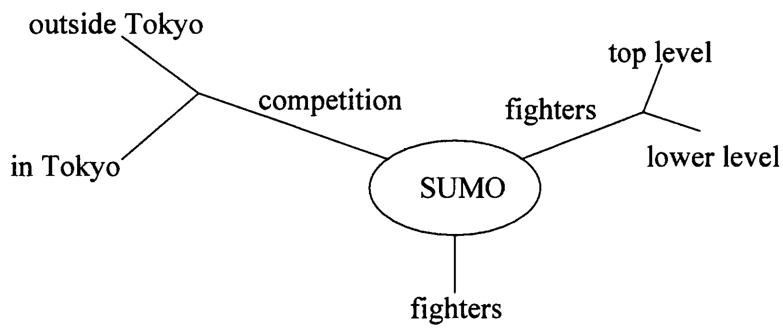

Figure 2.2. Example of semantic mapping (Nation, 2001, p. 129)

Nation (2001) argues that students need 'some clues' for semantic mapping to take place in L2 classrooms. These clues include the recall of a previously read story, a movie, or the learners' general knowledge of a topic (e.g., sumo = a Japanese style of wrestling). It is the interaction that "occurs during the building up of the semantic map that makes the activity contribute to vocabulary learning" (Nation, 2001, p. 129). Moreover, lexical features of words may often be difficult for students to agree upon, and semantic mapping activities promote a great deal of group interaction (Sökmen, 1997). These are essential features of semantic mapping in L2 classrooms, making a marked contrast with Sánchez's (2004) research study. In Sánchez's study, each student was given a copy of an expert net of *light* words. With the handout, the teacher and students "carried out an exhaustive analysis dealing with the relationships" (Sánchez, 2004, p. 93). Through such an activity, it is doubtful that students had active group interaction to determine the relationships between the words. In this regard, Sánchez's semantic mapping cannot be judged to be interaction-based L2 vocabulary learning. Instead, it is a teacher-oriented learning of lexical analysis. Accordingly, the good results of the experiment in the post-test had more to do with the participants' memory of the lexical relationships between tested words or other factors. As discussed earlier, they included learner awareness of *light* verbs after instruction and subsequent reinforcement through exposure to the media. Thus, it is not the case that the L2 lexical organisation of the participants in the experimental group was restructured into "an expert net" as the study claimed.

Lastly, I will briefly discuss the Pathfinder network algorithm (Pathfinder Analysis) as a data analysis and representation tool. Once data is elicited from participants in a psycholinguistic experiment, usually one of the following three multivariate data analyses is used to reveal the underlying organisational structure: dimensional representations (multidimensional scaling; MDS), tree constructions (cluster analysis) or Pathfinder Analysis. As Jonassen, Beissner and Yacci (1993) stated, each technique has advantages and disadvantages, and the choice of which is best depends on the data elicitation task a study adopts. In Sánchez (2004), the task was similarity ratings, and Pathfinder Analysis is the most commonly used technique for representing the similarity rating results. This is because Pathfinder "better represents local or pairwise comparisons between concepts in a knowledge domain" (Jonassen, et al., 1993, p. 74) than other techniques. However, Pathfinder is not good at analysing and representing the global information structure of data. If there are a number of links in the elicited data, the network representation would be too complicated to grasp the global relationship between them (Takeuchi & Utsugi, 1988). In other words, Pathfinder is powerful in revealing the structure of data which focuses on the local relationship among concepts. However, Pathfinder cannot detect significant differences between competing paths in the network because Pathfinder selects only the strongest association between nodes and all of the other relatively weaker ones are discarded. This is another reason why Pathfinder results need to be interpreted with care. This consideration was not given in Sánchez (2004), leading to another weakness of the study.

Despite the researcher's claim, using "an expert net" does not seem to have much promise when incorporated into L2 vocabulary instruction. Sánchez (2004) failed to provide convincing evidence for the claim that experimental instruction works well to structure native-like L2 lexical organisation, even though it appeared to be the case.

2.6 Discussion

In the previous sections, studies of L2 lexical organisation were reviewed. In this section, I will discuss three major issues that have emerged from the review from the viewpoint of investigating semantic clustering (i.e., cluster structure) of L2 lexical organisation. They are (a) research questions worth addressing, (b) the validity of longitudinal vs. cross-sectional studies, and (c) the merits and demerits of data elicitation techniques.

First, the literature review thus far has revealed that L1 lexical organisation is less varied than its L2 counterpart. This is what Meara and Schur (2002) have confirmed, supporting previous studies (e.g., Meara, 1978, 1983; Postman & Keppel, 1970; Riegel & Zivian, 1972; Szalay & Deese, 1978). This structural difference between L1 and L2 lexical organisation has been established by word association tests in some form or another. As indicated earlier, this is because in the past few decades, important findings in L2 lexical organisation research have mainly been made by word association tests that have included computer simulations of association. Accordingly, it would be very interesting to confirm whether this L1 and L2 variability difference can be generalised. To address this question, a different elicitation technique needs to be used in an attempt to find supporting evidence.

Meara and Schur (2002) provided another interesting finding: the tendency that L1 lexical organisation has more, smaller, non-connected components (clusters) than does L2 lexical organisation. Meara and Schur interpreted their results to mean that L1 speakers are more aware of the semantic relations between words and how the words may fall into distinct sets. It should be noted that the researchers gave a restricted word association test where participants selected only the most closely connected word among a limited set of 50. The task was sensitive to the degree of participants' awareness of links among possible pairs of tested words. Would the same results be produced by using a different elicitation task? Could another elicitation task tap into L1 and L2 participants' clustering behaviours and provide evidence for differences in awareness of lexical links? To my knowledge, no research has yet addressed these issues. This distinction between L1 and L2 lexical organisation merits further study.

Another interesting issue worth addressing is the arrangement of L1 and L2 lexical organisation. This is concerned with an observation Wilks and Meara (2002) made that "two networks with the same density could in fact be quite differently arranged in terms of how the connections between points are disposed and how many isolated points are present in the network" (Wilks & Meara, 2002, p. 319). It is possible for L2 speakers with high proficiency in the target language to have a large vocabulary, in-depth vocabulary knowledge, a good command of usage, and the same degree of lexical density as L1 speakers. But are their lexical organisations arranged differently from L1 speakers'? Does L2 lexical organisation still contain more disconnected, isolated lexical items than L1 organisation? These are important questions that have not been answered in any research to date. Thus, it was decided to investigate three main aspects of L1 and L2 lexical organisation in this research project: differences in variability, cluster size, and arrangement.

Second, it was requisite for me to decide whether the present project would adopt a longitudinal or cross-sectional approach in addressing the research questions identified above. A cross-sectional approach outweighs a longitudinal one in so far as the research purpose lies in revealing L1 and L2 differences in lexical organisation. This approach would involve giving a data elicitation task to both L1 and L2 groups simultaneously and analysing the results while searching for differences. Such a research design would reveal whether L1 and L2 organisational differences are actually present. Of course, as Haastrup and Henriksen (2000) did, it is worth conducting a longitudinal study when the question is whether L2 learners develop a native-like lexical organisation by the effect of some treatment or after months or years of EFL/ESL learning. However, as pointed out above, it is extremely difficult to detect a change in participants' lexical organisation. Haastrup and Henriksen found only two participants out of the original 17 participants in their study who managed to show at least some change over the years. It is certainly true that restructuring L2 lexical organisation is "both a difficult as well as slow process" (Haastrup & Henriksen, 2000, p. 227). This is also the case with L1 lexical network building, which is widely accepted by many researchers (see Aitchison, 2003, for a comprehensive overview). It is not easy to detect progress of a phenomenon which is slow and difficult in nature. Also, it seems futile to conduct a study attempting to prove that a process is slow and difficult.

I criticised the seemingly promising results of instruction using "an expert net" that Sánchez (2004) reported. It is doubtful that having L2 learners study an example of an L1 lexical network really brought about an approximation to native-like lexical organisation. As argued in the previous section, the results could have been attributed to the good memory of L2 learners of a limited set of target words or other factors. This is not an approximation of L2 lexical organisation into a native-like lexical network structure. All in all, it has little pedagogical significance to have L2 learners learn a native-like lexical network within such a short period of time of only three months. Sánchez's L2 participants might have learned or memorised the organisation of the words during this period. Answering the question of whether L2 lexical organisation is actually different from L1 lexical organisation is more crucial.

Third, the literature review thus far has revealed some merits and demerits of data elicitation techniques for investigating lexical organisation. My specific interest lies in the cluster structures of lexical organisation. In this regard, word association tests appear to be weak in validity. The reason is simple: what participants are required to do in a WAT-based experiment is to associate a single word with another that participants think is most strongly linked with it. A WAT draws attention to an associative relationship between two words among many others. This holds with other types of WATs where, for example, participants are directed to find the second, third ... or *n*th word as well that is associated with a stimulus word in a WAT. Every time a participant is pondering the *n*th associated word to the stimulus, she is forced to find a pair of associated words. The data this multiple WAT produces is often only a short association chain (Kruse, Pankhurst & Sharwood Smith, 1987; Wilks, 1999). Thus a WAT does not seem to be suitable to tap into the underlying cluster structure of lexical organisation.

As stated above, most of the important findings on L1 and L2 lexical organisation have been made by word association tests. This is also true with the recent studies of Meara and Schur (2002), Schur (2007), Wilks and Meara (2002), and Wilks, Meara and Wolter (2005) reviewed earlier. However, the present research project is taking a different approach to lexical organisation. It will examine how L1 and L2 lexical organisations are clustered and whether they are different from each other. My interest does not lie in the associations of two words, but in the clustering of a group of words by participants. In reality, some lexical items in the mental lexicon might be isolated from other words, some are linked with only one word, and others are clustered with two, three ... or *n* words. This is the aspect of lexical organisation to be explored, while taking into consideration the findings and unanswered questions of previous studies.

Similarity ratings are not an entirely suitable methodology either. I argued that the reliability of similarity ratings is not as high as generally assumed. Similarity rating tasks have participants rate the degree of similarity between every pair of tested words. Sánchez (2004) asked participants for 171 ratings on the relationships between 19 *light* verbs. There were too many trials for a limited set of 19 words for participants. The task demanded more effort and concentration than normal. Participants likely lost interest and got tired of it, which calls task reliability into serious question.

Given the unsuitability of word association tests and similarity ratings for researching the cluster structures of lexical organisation, I decided on a sorting task. A sorting task directs participants to think over how words can be grouped together in terms of relatedness or similarity. It is reasonable to assume that how words as a cluster in lexical organisation are arranged will be reflected in the results. However, it should be noted that the sorting tasks Haastrup and Henriksen (2000) developed did not succeed in tapping into the clustering behaviours of participants. This is because the results were heavily affected by whether participants knew the meanings of the words being tested. Their sorting tasks contained low frequency adjectives among the tested words, and many participants did not know their meanings even at later stages of the longitudinal study. Accordingly, the task results reflected participants' lexical knowledge (i.e., whether they knew the meaning of the words) more than how well they had developed L2 lexical organisation in the mental lexicon. For this reason, it was felt that a different kind of sorting task needed to be developed.

In the following sections, I will review four L1 lexical organisation studies and compare the data elicitation tasks used as well as other facets of them, including how data was analysed. I will start with Miller (1969).

2.7 Miller (1969)

2.7.1 Summary

Miller (1969), a pioneering study in psycholinguistics, addressed how lexical knowledge is organised and stored in L1 memory. Miller adopted a sorting task using 48 nouns. They were 24 names of objects (e.g., *anchor, bleach, yacht*) and 24 names of non-objects (e.g., *aid, battle, counsel*). Each word was typed on a card. On each card there was a short definition specifying the sense of the noun and a simple sentence illustrating the use. 50 students at two universities in the US, all L1 English, participated in the experiment. They were individually directed to sort the 48 nouns into piles on the basis of similarity of meaning. There was no restriction on the number of piles and the size (how many cards a participant could put into a pile). Most participants spent from five to 30 minutes to complete the task. They made 14.3 piles on average (SD = 5.0).

The data was put into an incidence (i.e., co-occurrence) matrix. The matrix data was then submitted to two types of cluster analysis, the connectedness method and the diameter method. Miller attempted to prove that the two solutions would give more or less the same answer. It was revealed that "about 70% of the clusters indicated by the two methods are common to both" (p. 181). Miller (1969) thus concluded that cluster analysis is a valid technique to analyse a group matrix of sorting data.

Using the diameter method dendrogram (i.e., tree graph), Miller (1969) confirmed five basic verbal concepts: names of living things (e.g., *mother, cook, doctor*), names of non-living things (e.g., *tree, plant, root*), quantitative terms (e.g., *inch, measure, number*), kinds of social interaction (e.g., *battle, kill, deal*), and psychological terms (e.g., *thrill, ease, fear*). To test the hypothesis that participants did the sorting on the basis of shared conceptual features, all 48 nouns in the experiment were looked up in Roget's *Thesaurus*, and a maximum number of shared categories was tabulated for each pair of lexical items. There were 1,128 pairs. The mean proximity for pairs of items was plotted as a function of the number of shared *Thesaurus* features. It was found that "there is a rough correlation, although the variability is so great that no precise prediction of the data could be derived from Roget's classification" (p. 183).

Miller argued that one would expect the sorting task "to be carried out largely on the basis of what the nouns presuppose, rather than what they assert" (p. 185). In that case the results would reveal the degrees of compatibility among presuppositions, and the structure that emerged should be a presuppositional structure. Eventually, with the sorting task data, this presuppositional structure turned out to be hierarchical. His claim was confirmed, as in the example of the hierarchical structure *knight* < *man* < *person* < *being*. That is, "*being* dominates *person*, since *person* is undefined for non-living things; *person* dominates *men*, since *man* is undefined for non-persons; and *man* dominates KNIGHT, since KNIGHT is undefined for non-men" (Miller, 1969, p. 185). Miller also confirmed that the results obtained from the sorting task had linguistic relevance. For example, the first five nouns (*mother, cook, doctor, umpire* and *knight*) in the dendrograms were linguistically related to each other. This offered evidence for the presupposition-assertion hierarchical topology Miller claimed.

2.7.2 Commentary

Although Miller's study was published in 1969, it continues to be of utmost value to current research and should be closely studied and used. His rigid approach to lexical organisation is particularly noteworthy. However, it had one methodological flaw: giving a definition and example sentence for each word in the experiment failed to achieve the study's purpose. That is, Miller's attempt to specify the meaning of the nouns used in the sorting task was problematic. In the study, by giving definitions and example sentences, participants were directed to work with only one specific meaning of each word and think over the cluster relationship of each word with other words. Forty-eight nouns had been chosen in advance in accordance with the criterion that the words should be grouped into one of five clusters, including names of living things, names of non-living things, and quantitative terms.

The problem lies in exclusively deciding the meaning of a word by giving a definition and example sentence and having participants work only with this one meaning during task completion. Actually, some nouns were not sorted by the presupposed criterion Miller (1969) established, but by idiosyncratic decisions participants made. Miller noticed the problem from the fact that *fish* was clustered with *plant* and *root,* and *wheel* with *jack.* He stated that an unsuitable definition or a sentence example contrary to the presupposed meanings affected sorting behaviours. He exemplified the shortcoming with a confusing definition of "an *animal* that lives in water and breathes with gills" for *fish* and with the sentence example "most cars come with a jack fitted in the trunk component" for *jack.* Miller admitted that the sentence example for *jack* led many participants to connect it with *wheel* "not on presuppositional grounds, but because a *jack* is used to raise a car when removing a *wheel*" (p. 187). On the face of it, it might have been possible to tune the definition and sentence example more finely. However, it is not possible to make all definitions and sentence examples exactly reflect the presupposed meanings a researcher is interested in. Also, there will certainly be cases where participants reveal idiosyncratic clustering behaviours as those in Miller's study did. This will be inevitable, no matter what precautions are taken. Thus, a sorting task would work better by giving participants only tested lexical items without definitions and sentence examples. The results would reveal how participants sort words into groups, free from predetermined definitions of word meanings. The results would tell us more about how L1 and L2 lexical organisations are differently structured and how they could be reliably compared to each other.

More importantly, Miller (1969) confirmed that sorting tasks are a valid method for investigating cognitive structures. A distinctive merit of sorting tasks is their high content validity, meaning "the representativeness of our measurement regarding the phenomenon about which we want information" (Mackey & Gass, 2005, p. 107). Miller (1969) has high content validity. Sorting tasks, or free sorting tasks, direct participants to put lexical items into groups according to the way they think that the task items are related to each other. Moreover, in free sorting, participants sort words into as many groups as they wish, with any number of words placed in a group. Tasks are expected to reveal the way lexical items are organised and stored in the mental lexicon. Thus sorting tasks have high content validity for answering how "lexical information is subjectively organized and stored [in L1 lexical organisation]" (Miller, 1969, p. 169).

Besides the free sorting task Miller (1969) adopted, there are three other types of sorting tasks: *fixed-sorting* (when the number of the categories is fixed in advance), *graded-sorting* (which occurs when the categories are required to be in rank order), and *multiple sorting* (which occurs when the participant makes more than one sorting of the same set of objects) (Coxon, 1999). Given these various types of sorting tasks, however, it has been noted that free sorting has been predominantly used in studies of cognitive structures and verbal behaviours. This simple fact appears to add to the validity of free sorting in examining lexical organisation. Studies of cognitive and lexical structures using sorting tasks started in the 1960's. Almost all were done on L1 cognitive and lexical structures (e.g., Anglin, 1970; Burton, 1975; Chaffin & Herrmann, 1984; Fillenbaum & Rapoport, 1971; Miller, 1967, 1969; Preece, 1976; Rapoport & Fillenbaum, 1972; Rough, 1994; Takane, 1980). Evidently, there have been no studies of L2 lexical organisation using sorting tasks until now, except Haastrup and Henriksen (2000) and their preliminary studies (Haastrup & Henriksen, 1998; Henriksen & Haastrup, 1998). As discussed earlier in this chapter, L2 lexical organisation studies have been done extensively using word association tests, and they have yielded important findings. Considering this fact and the merits of the sorting tasks identified above, it can be argued that a sorting task is a valid data elicitation technique to address issues of lexical organisation.

2.8 Rapoport and Fillenbaum (1972)

2.8.1 Summary

This study is based on Lyons' (1968) claim that meaning should be treated as a function of the relationships between the meanings of words. To address it, Rapoport and Fillenbaum (1972) conducted two experiments, the first with colour names and the second with the HAVE family of verbs. I will review the latter experiment which used 29 HAVE verbs (e.g., *accept, earn, hold, take*): the first experiment will be referred to as needed.

Rapoport and Fillenbaum (1972) explored an issue regarding the sort of semantic structure required to accommodate similarity judgements for a set of HAVE verbs. They claimed that "in contrast to colour names [which they investigated in their first experiment], these terms constitute a semantic field whose boundaries are rather ill-defined and indefinite" (Rapoport & Fillenbaum, 1972, p. 94). The study had 58 participants, who were all undergraduates at an American university. They were divided into Group HTM (HAVE verb Tree Construction Task, 17 students) and Group HD (HAVE verb Direct Sorting Task, 41 students).

Participants in Group HTM were given options for the task. They were given a list of 29 HAVE verbs and a blank piece of paper. From the list they picked the two words they thought were most similar to each other. They wrote the pair on the paper and connected them with a line, labelling it "1". Then there were two options participants could choose from for the remaining words in the list. One option was to pick the word which they thought was the most similar to either of the two words they had already selected. They wrote it down on the paper and connected it to the appropriate word already selected, and labelled the connecting line "2". The other option was to decide that two of the remaining words were more similar to each other than either of them was to either of the two words already selected and joined together. They wrote them down on the paper, and connected the two new words with a line and labelled it "2". As the experiment went on, participants might have had a third option, which was to connect any two of the trees (i.e., linked group of words) together. If they found two words on two separate trees that were more similar to each other than any other word on the remaining list was to any other word on the tree, they connected the two words. They labelled the connecting line according to the sequence already started. Participants continued the task until all words were used and until they had connected all separate trees into one tree. Participants were allowed up to two hours to complete the task, but most of them completed it in less than one hour.

Group HD was given a sorting task. They were directed to sort the same 29 HAVE verbs in terms of similarity of meaning. Each word was printed on a card, and the deck of cards was arranged in alphabetical order. Participants laid out the cards and looked carefully over all the words. They put the words into piles on the basis of similarity of meaning and were allowed to make as many piles as they wanted and have as many or as few words as they liked in any pile. When they completed these steps, participants looked over their piles and made any adjustments or changes they felt appropriate. On average, participants took between 10 to 15 minutes to sort the words in the deck.

Cluster analysis confirmed that the participants in Group HTM constructed six sub-trees, including a NEED cluster [*lack, need, want*], a RECEIVE cluster [*buy, accept, find, earn, gain,* *get, receive*] and a TAKE cluster [*steal, take*]. Although the composition was not always the same, the participants in Group HD also made six sub-trees, such as a NEED cluster [*lack, need, want*], a RECEIVE-TAKE cluster [*accept, get, receive, buy, earn, gain, find, steal, take*] and a GIVE cluster [*lose, lend, give, offer, get rid of, sell*]. Rapoport and Fillenbaum (1972) argued that the results of Group HTM and Group HD were "closely related" (p. 120). They pointed out that four clusters were exactly the same between the two groups.

Multidimensional Scaling (MDS) analysis was then done on the Group HTM data. Rapoport and Fillenbaum (1972) stated that the two-dimensional MDS representation was difficult to interpret. Tight clusters appeared when plotting the data points and any attempt at MDS dimensional representation imposed structure on the data which was not actually there. The researchers argued that MDS representation "can easily lead one to overlook special features of the configuration" (Rapport & Fillenbaum, 1972, p. 121). Further analysis using a three-dimensional MDS solution yielded similar discouraging results. On the other hand, MDS analysis on the Group HD data was easier to interpret. The horizontal axis was the OFFER-RECEIVE dimension. On the left-hand side of the axis there were words that indicated a purposeful activity. On the right side of the axis there were words that implied passive reception. Given these mixed results together, Rapoport and Fillenbaum concluded that doing MDS analysis on the HAVE verb data was not appropriate. Axes of the MDS solutions were difficult to interpret, and the terms did not distribute in an "intuitively sensible way in the space" (p. 127).

Rapoport and Fillenbaum (1972) claimed that a spatial structure would provide the most appropriate representation of the colour term data, while in regard to the HAVE verbs, there were not such strong grounds to use a spatial structure as in the case of colour names. It was more appropriate to hold "somewhat weaker grounds for believing that a looser sort of taxonomic structure would provide a better representation" (p. 129). Rapoport and Fillenbaum came to the conclusion that cluster analysis yielded sensible representation of the HAVE verb data and allowed them to reasonably infer that the underlying structure was hierarchical.

2.8.2 Commentary

This study addressed the question of how L1 semantic structures are organised. It was carefully designed, using different data elicitation techniques (i.e., tree construction and sorting tasks) for the purpose of the study. It compared the sensitivity of MDS and cluster analysis in revealing underlying structures. Their analyses were solid and the interpretation of the results was reasonable. Keeping these merits of Rapoport and Fillenbaum (1972) in mind, I will discuss the problems of the reliable elicitation task and valid data analysis done in the study.

Regarding the cluster analysis results, Rapoport and Fillenbaum (1972) claimed that HTM (the tree construction task using HAVE verbs) and HD (the sorting task using HAVE verbs) were closely related and yielded essentially the same results. However, this claim should be called into question. It is true that both participant groups generated, on average, the same sub-trees in number, and four of them were more or less identical in composition. However, when taking a closer look at the other sub-trees generated, they were distinctly different from each other. Group HTM produced a RECEIVE cluster [*buy, accept, find, earn, gain, get, receive*] and a TAKE cluster [*steal, take*], whereas Group HD generated a BRING-RETURN cluster [*bring, return*] and a RECEIVE-TAKE cluster [*accept, get, receive, buy, earn, gain, find, steal, take*]. Thus, the configurations of the two task results cannot be regarded as being identical at all. It also should be noted that in Group HTM, the words *bring, lose* and *use* formed separate clusters, and they were not components of any of the sub-trees. The researchers' interpretation that HTM and HD produced essentially the same results deviated from the actual results.

A more serious problem arose from the question of comparing the results of totally different data elicitation techniques, tree construction and sorting tasks. As discussed earlier, the tree construction method was an extremely demanding task that required participants to judge the degree of similarity between every possible pair of words, and then to connect and label them in ranked order. In this sense, the task was similar to similarity ratings (see the review of Sánchez (2004) in section 2.5). The task was also very demanding and forced participants to work harder than required in a standard psycholinguistic experiment. Moreover, participants were allowed to take up to two hours to complete the task. This is a long time for participants to continuously concentrate on a demanding task. Meanwhile, in a sorting task, or a free sorting task, participants simply sorted tested words into groups of meanings they thought were related to each other. There were no restrictions on the number of word clusters and the number of words per cluster. Compared with tree construction, the sorting task was far less demanding and less time-consuming (participants completed the sorting task, on average, within 10 to 15 minutes). Thus, the two data elicitation techniques Rapoport and Fillenbaum used were markedly different from each other and likely tapped into different facets of lexical organisation.

However, it was a good decision to give two different tasks to two different participant groups, as Rapoport and Fillenbaum did. Ideally, to test whether two tasks are tapping into the same cognitive domain, they should be given to the same participant group on different occasions. In this way, the results could be analysed based on the degree of a within-group correlation between two data sets. However, in reality, it is difficult to conduct two experiments with one participant group while keeping the same level of reliability. Two of the three data elicitation techniques in Rapoport and Fillenbaum (1972) have been discussed in this section. The third one, namely the method of complete undirected graph making, has not. In short, the third task asked participants to decide which was the most similar pair of words in a list and which was the next most similar pair. It should be noted how complicated the task directions were: "Work slowly and carefully; this is a difficult task; take your time" (Rapoport & Fillenbaum, 1972, p. 97). Because of this, the third task of complete undirected graph making could be considered extremely demanding and complex for participants to carry out consistently. Thus, the task appears to be the least suitable in gaining consistent within-subject and between-subject reliability. Rapoport and Fillenbaum did not explicitly state their conclusion concerning which task was the most valid and reliable. But, given the comparison of the tasks above, the method of complete undirected graph making was the least reliable and the sorting task was the most promising for tapping into lexical organisation.

Rapoport and Fillenbaum (1972) uncovered integral facets of L1 semantic structures, which should be useful in L2 lexical organisation studies as well. Given the superiority of sorting tasks over other elicitation tasks, there is certainly great potential in examining lexical organisation using sorting tasks followed by valid statistical analyses.

2.9 Preece (1976)

2.9.1 Summary

Preece (1976) attempted to confirm that three kinds of data elicitation tasks would produce the same patterns of relationships between a set of mechanics concept words. He claimed that mechanics concepts are useful in investigating cognitive structure because their scientific meanings can be simply defined by means of basic concepts, e.g., *area, volume, density, velocity* and *acceleration*. Preece proposed a model, namely the digraph model, in which the semantic relationships between words could be neatly represented.

The mechanics concepts used in the research were 15 terms such as *weight, force, energy, distance,* and *volume*. The participants were 28 university science graduates at a university in England. First, they took a free word association test in which they were told to write down the first word that came to mind when seeing a mechanics concept term. Three weeks later, they took a controlled WAT in which they were instructed to write down the name of the physical quantity which they considered to be the most closely associated with a given mechanics concept term. After a further interval of at least one week, they did a tree-construction test (the same task as Rapoport and Fillenbaum (1972) adopted) while carefully working with a list of mechanics concepts and were instructed to construct a tree of relationships among the concepts.

The group average half-matrices of inter-concept proximities for word association tests and separations for the tree-construction test were created. Cell entries in the half-matrices were simply calculated as the mean values of the corresponding entries in the individual half-matrices, and analysis was confined to only these group matrices. Preece (1976) did this analysis based on Deese's (1965) argument: meaning is "a property of a culture, not an individual, consequently he [participant] defined associative structure in terms of cultural norms of word association" (Preece, 1976, p. 5). Preece also argued for adding up individual sets of data. He stated that "separations in a participant's tree are not likely to be even monotonically related to the corresponding distances in his cognitive structure, but that pooling individual data will produce more valid results" (p. 5).

Data collected by the three data elicitation techniques were submitted to Waern's (1972) method of analysis. Preece claimed that the method "yields a graph depicting the concepts as points and proximities as lines, and it was proposed as an alternative to dimensional methods [e.g., MDS, cluster analysis, factor analysis] for the investigation of cognitive structure" (Preece, 1976, p. 4). The graphic representation succeeded in producing a meaningful structure quite similar to the digraph model. However, Preece admitted that the method was "extremely simple and ignores much of the metric information available" (p. 6).

High correlations among the three tasks were achieved (.831-.996). Thus, the tasks revealed a similar underlying structure. It was shown that the controlled WAT data were closer to the digraph model of the mechanics concepts (r = -.752 for unweighted controlled WAT and r = -.774 for weighted controlled WAT) than others. Preece stated that the higher correlations between the controlled WAT data and the model arose because the controlled test limited responses to physical quantities. The free WAT generated "a more general associative structure somewhat less well described by the physics model" (Preece, 1976, p. 6).

Preece claimed that all three tests revealed a similar underlying structure and provided "validation of the tests as means of mapping cognitive structure" (Preece, 1976, p. 7). It was concluded that what the three cognitive structure tests revealed was well represented by the digraph model the research proposed. Preece reported that the controlled WAT was a particularly valuable method for mapping the cognitive structure of a specific semantic domain.

2.9.2 Commentary

Preece (1976) isolates important features that should be attended to in L2 lexical organisation studies. Among them, two will be discussed: (a) testing a model of cognitive structure with experimental data and (b) the relationship between individual differences and group data.

First, Preece (1976) had an advantage over other studies in that he dealt exclusively with mechanics concepts. These are terms that we would not expect ordinary people to know, though they should be familiar to graduate students of science. Because of the limited scope of his research, it was possible for him to postulate a model of the semantic structure of the data. In this sense, Preece (1976) is in line with Rapoport and Fillenbaum (1972), which was reviewed in the previous section. They also posed the question of whether Lyons' (1968) model of meanings would actually function in human memory as had been claimed. It appears that focusing on a limited domain of cognitive structure made it possible for both studies to conduct their model testing-based research, but this is usually not the case with L1 and L2 lexical organisation research. This is because lexical organisation in the mental lexicon is composed of a far larger number of components (e.g., 3,000, 5,000, 10,000, or 200,000 words). It is not possible to postulate a model of lexical organisation duplicating all the relationships in the way that Preece (1976) and Rapoport and Fillenbaum (1972) did. Their work highlighted an inherent limitation in conducting L2 lexical organisation research.

Second, it should be noted that Preece (1976) provided a solution to the issue of the relationship between individual differences and group data. The digraph analysis Preece adopted was done on group half-matrices that were simply taken as the mean values of the entries in the individual half-matrices. The results were primarily concerned with the group data of participants. However, as the researcher argued, meaning and its structure in memory is not confined to an individual but rather the property of a culture. Preece also argued that "pooling individual data will produce more valid results" (p. 5). Eventually, group data analysis is not only concerned with group data but also with individual participants. This is similarly the case with other multivariate data analyses such as factor analysis, cluster analysis, and multidimensional scaling, which involve themselves in group data analysis.

It is not clear whether the graphic analysis Preece (1976) adopted was the most valid one for the data the study collected. This is due to the fact that the analysis "ignores much of the metric information available" (p. 6) to produce a meaningful and perceivable graphic representation of a mechanics concept relationship. The graphic representation the analysis made oversimplified underlying cognitive structure by collapsing much of the information gained in the experiment. In this regard, multidimensional scaling (MDS), which similarly processes this kind of space-oriented data while also keeping the information available, seems to have more validity. In spite of this flaw, Preece (1976) exhibited robust research design, related a model to empirical data in the area of research, and provided in-depth data analysis.

2.10 Routh (1994)

2.10.1 Summary

This study focused on how quantifiers (e.g., *all, several, few*) are established in L1 cognitive structure. He specifically analysed the results of a sorting task that was designed to reflect the underlying semantic structure of quantifiers. Routh thoroughly reviewed studies of quantifiers (Borges & Sawyers, 1974; Moxey & Sandford, 1992, 1993; Newstead, Pollard & Riezebos, 1987). Then he claimed that the relevant variables required for an adequate profile of quantifier meaning had not been identified yet. He argued that the elicited knowledge of quantifiers should be examined and represented by means of two multivariate data analyses: multidimensional scaling (MDS) and cluster analysis. Routh was also interested in which of the two statistical analyses would work better to reveal the underlying semantic structure of quantifiers.

Routh (1994) also attempted to examine the influence of "the set size being referred to" (p. 204) on the meanings of quantifiers. That is, the study tested whether participants working with quantified phrases that were fixed at one set size would produce different results from those working with quantified phrases fixed at another set size. For example, a participant group sorting quantified phrases of *all (none, each, half. . .) of the students in a class of 12* may or may not produce results different from another group sorting quantified phrases of *all (none, each, half. . .) of the students in a class of 24.*

Participating in the experiment were 129 undergraduates at a university in England. They were in a first-year practical class in Psychology. As materials for a sorting task, 20 exemplars of quantified phrases were prepared in the following pattern: *Q of the Students in a Class of N.* Within each phrase, Q was replaced by one of 20 quantifiers. Routh prepared four sets of phrases. Within each set, *N* was fixed at one of the four possible values of set size, i.e., *12, 24, 48,* and *96.* In the experiment, participants were randomly assigned to one of the four groups. Each participant was given a relevant, randomised set of 20 typed slips of paper sized 6 × 5 cm., with each having a different exemplar of a quantified phrase. Participants were directed to sort them into a freely chosen number of clusters on the basis of similarity of meaning.

As a preliminary analysis, a co-occurrence matrix was constructed for each set size. The matrix revealed very little variation as a function of set size. Routh (1994) concluded that the estimated structures were essentially invariant over set size. Accordingly, Routh conducted analyses on a new matrix for the data pooled from participants. The results of a two-dimensional MDS were given, showing the quantifiers as points in a plane. Routh stated that it was "very clear that the points cannot be projected on to an obvious global dimension corresponding to quantity (whether of amount or proportion): However, some global aspects of quantifier meaning are revealed" (p. 207). There was a left-to-right distinction between the 'fuzzier' quantifiers (e.g. *few, several, many, most*) and the more precise ones (e.g. *all, none, each*). It was also revealed that there was a vertical contrast between relatively small and relatively large quantities. Furthermore, the results of cluster analysis were shown as a dendrogram. In the representation, the quantifiers were represented by terminal nodes of the tree on the right-hand side. Routh argued that most quantifiers were well defined in terms of unique features, being combined into clusters that were fairly interpretable. He borrowed labels for the clusters from Quirk, Greenbaum, Leech, and Svartvik (1985). The clusters appeared to be consistent with the categories of indefinite pronouns derived by using traditional techniques. The major clusters were ordered from the top to the bottom of the dendrogram with more precise meanings at the top and fuzzier meanings at the bottom. The vertical arrangement of the clusters approximately corresponded to an ordering in terms of quantity. Routh argued that the dendrogram neatly captured the semantic structure of quantifiers in the mind.

Given the comparison above, Routh (1994) concluded that cluster analysis is superior to MDS in revealing the nature of the common and distinctive features involved in quantifier meanings.

2.10.2 Commentary

Routh (1994) was a study of semantics focusing on the domain of quantifiers, and the research questions posed in it owed much to the traditions of psychology. The paper demonstrated some aspects of solid research. Among them, three will be discussed: (a) the ability of the *Q of the Students in a Class of N* frame in tapping into the organisational structure of quantifiers, (b) sorting tasks as a data elicitation technique, and (c) MDS and cluster analysis for examining cognitive structures. The limitations of the study will also be discussed.

First, it was wise for Routh (1994) to use the *Q of the Students in a Class of N* frame instead of investigating the relationship between isolated quantifiers. Using this frame, participants in the experiment readily worked out how quantifiers are used, related them to other quantifiers, and sorted them into clusters. Eventually, it was found that the set size parameters brought about no discriminating differences among participants. Because of this finding, the subsequent analyses Routh could be considered reliable.

However, the frame Routh (1994) used in addressing the underlying lexical organisation of content words need not be adopted. Content words (e.g., nouns, verbs, and adjectives) can stand alone and be meaningful, while non-content words (quantifiers in the study) cannot. Establishing a frame to specify the meanings of content words in an experiment does not help to gain reliable information. For example, even though the meanings of nouns can be specified by short definitions or sentence examples, they often do not function well. The shortcomings of the definitions and sentence examples in Miller (1969) should be noted. One problem, among others, lay in the difficulty of offering a clear-cut single definition or sentence example for each content word. The definitions and examples often led participants to unpredicted behaviours in the experiment, evoking idiosyncratic responses. In this sense, it was valid that Rapoport and Fillenbaum (1972) used *colour* names and HAVE verbs as isolated words, as did Preece (1976) with mechanics concept nouns.

Second, as Routh (1994) showed, sorting is a valid task to address the underlying semantic structure of quantifiers in lexical organisation. To elicit structural organisation in cognition and perception, there are three major data elicitation techniques: word association tests, similarity ratings, and sorting tasks (Jonassen, Beissner & Yacci, 1993). In L2 lexical organisation studies, word association tests have been used extensively. Particularly in the past decade, often with computer simulations, WATs have been used in important studies such as Fitzpatrick (2006, 2007, 2008), Meara (2004, 2006, 2007a, 2007b), Meara and Schur (2002), Nissen and Henriksen (2006), Wilks and Meara (2002), Wilks, Meara and Wolter (2005) and Yokokawa, Yabuuchi, Kadota, Nakanishi and Noro (2002). On the other hand, sorting tasks have rarely been used in L2 lexical organisation studies, except in a series of studies conducted by Haastrup and Henriksen (1998, 2000) and Henriksen and Haastrup (1998). Routh (1994) compared sorting tasks to similarity ratings, arguing that when "a large set of objects must be judged, the more traditional method of pairwise comparisons [similarity ratings] can make prohibitive demands on a respondent's time and motivation. By comparison, the freesort method enables a respondent to make judgements about an entire set of objects fairly efficiently" (p. 203). The negative effect of similarity ratings was identified earlier in the review of Sánchez (2004). The study had participants judge the degree of similarity of 171 pairs of *light* words. However, in sorting tasks, participants can make judgements regarding a whole set of tested objects within a relatively short time. This is another merit of sorting tasks. Sorting tasks also have high face validity for tapping into latent clustering structures by asking participants to make clusters of words that they think can be grouped together by meaning.

Third, Routh (1994) made a fair comparison of MDS and cluster analysis. He then confirmed that cluster analysis is a better technique than MDS in addressing the issues of the organisational structures of quantifiers. As Johnson and Tversky (1984) stated, the results of similarity judgments are best represented and explained by tree model statistics of cluster analysis. Among multivariate data analyses (e.g., MDS, factor analysis, principal component analysis and multiple regression analysis), cluster analysis can be distinctively characterised by its calculation procedure. Simply speaking, it puts individuals or groups into clusters in view of the distance between tested words. Thus, results reveal structures in that "objects in the same cluster are more similar to one another than they are to objects in other clusters" (Hair, Black, Babin, Anderson & Tatham, 2006, p. 554). Accordingly, the underlying structures of sorting task results are represented succinctly by cluster analyses.

Taking all of this into account, there is certainly no single multivariate data analysis technique that can be applied to every kind of data obtained from a psycholinguistic experiment. The type of analysis to be used should be decided by taking into account the nature of the goal and the domains to be investigated. However, as Routh (1994) concluded, considering the degree of similarities in cognition when using sorting tasks, cluster analysis is the most valid type of statistical analysis.

Routh (1994) might have been more convincing if the results had been submitted to a post hoc analysis to determine how many clusters as a final cluster solution there should be. By conducting such an analysis, the relationship between the 20 quantifiers would have been much clearer. However, I feel that Routh (1994) confirmed that sorting tasks are an appropriate approach to many areas of research, including L1 and L2 lexical organisation studies. The paper also showed that for data collected by sorting tasks, cluster analysis is the most effective method for revealing the underlying organisational structure.

2.11 Discussion

In sections 2.7 to 2.10, several studies of L1 cognitive structures were reviewed. All were robust and carried out with theory-based research questions, regardless of the domain. In section 2.6, the L2 lexical organisation studies reviewed in the former part of this literature review were discussed. It was concluded that in investigating clustering behaviours of L1 and L2 participants and the underlying lexical organisation, sorting tasks are the most promising. However, the sorting tasks that Haastrup and Henriksen (2000) developed, which were basically the only sorting tasks in L2 lexical organisation studies up until then, had more to do with the participants' lexical knowledge. The tasks did not have enough sensitivity to examine lexical organisation and were inappropriate for the purpose of the present research project. Thus, I argued for the importance of finding another sorting task. In this section, I start the discussion with the sorting tasks used in L1 cognitive structure research followed by a discussion of the statistical analyses of them. Lastly, I'll address the issue of group data analysis and individual differences.

First, it should be noted that the sorting tasks used in the L1 cognitive structure research reviewed (Miller, 1969; Rapoport & Fillenbaum, 1972; Routh, 1994) were always free sorting tasks. They did not employ any other types of sorting tasks, such as fixed sorting, graded sorting or multiple sorting tasks. In contrast with a free sorting task, these tasks restricted participants in some way or another: the number of the categories they could make (fixed sorting), the rank-order of the results they imposed (graded sorting) and the number of times they could do the task (multiple sorting). Thus each sorting task should be selected depending on the specific research purpose.

Why, then, did Miller (1969), Rapoport and Fillenbaum (1972) and Routh (1994) all employ free sorting tasks? Broadly speaking, they were all interested in the domain of L1 lexical organisation and addressed the issue of how it is arranged. Because of this, the researchers did not want to impose any sorting restrictions on participants. If they had done so, it would have deviated from their research questions. Participants would not have sorted words freely, and the results would not have reliably reflected the underlying intact structures. By having participants sort words into groups without restriction, the results were expected to reveal the underlying structures as they are. In this sense, all three studies shared a common research principle, and free sorting tasks met the expectation.

As argued in section 2.6, word association tests are not suitable to examine how L1 and L2 lexical organisations are structured by means of clustering behaviours. WATs are primarily concerned with the associative relationship between two words. The directions were as follows: "Decide which one of the verbs in the Verb Box best completes the pair for each verb in that list" (Meara and Schur, 2002); "Put down next to number 1 the name of the physical quantity which you consider to be most closely connected with the given quantity" (Preece, 1976); and "Read each set of words and circle any words in a set that you consider to be associated" (Wilks and Meara, 2002). This kind of WAT, in which participants elicit stimulus-response pairs, is common in the most recent WAT-based research. Note, for example, Fitzpatrick's (2006) "Please write down the first word you think of when you read each of the words" and Racine's (2008) "Read the following list of words and write down the first English words that come to mind." It is widely confirmed that word association tests have produced some important findings on facets of L1 and L2 lexical organisation, but having participants identify stimulus-response pairs does not fit well with the present research of addressing the cluster structures of lexical organisation. A data elicitation technique should directly tap into lexical clusters (i.e., groups of words) in lexical organisation. Thus, a sorting task is preferable to a word association test and is also more valid.

Sorting tasks have been used in numerous studies, including studies of cognitive structure, person perception, conceptions of psychopathology and child cognitive development. See Coxon (1999), Jonassen, et al. (1993) and Rosenberg (1982) for extensive reviews of sorting task studies. It is noted that these are all related to the cognitive domains, where relatedness (similarity) of concepts, perceptions and knowledge play a significant role. Thus it is natural that the underlying structures be efficiently examined and represented by sorting tasks. L2 lexical organisation is also within this category. As stated earlier, Miller (1969) and Rapoport and Fillenbaum (1972) adopted free sorting tasks. They argued that a free sorting task is easy to implement, requires less time and places less fatigue on participants compared with other demanding tasks. This merit is in line with Routh (1994), who states "the freesort method enables a respondent to make judgements about an entire set of objects fairly efficiently" (p. 203). Considering the lack of sorting tasks in L2 lexical organisation studies, these tasks are useful techniques to tap into structural knowledge.

Second, the literature review showed that it is necessary to run multivariate data analysis in examining how lexical organisation is structured. Pathfinder analysis, which Sánchez (2004) employed, could be a candidate, but I would argue against it mainly because results are not easy to interpret and require subjective judgements on the part of researchers. That is, the analysis requires "further interpretation [on the part of a researcher] because the links in the graph are not labeled or differentiated semantically" (Cooke, 1990, p. 228). Thus, the distinction between the links might be ambiguous and the interpretation might turn out to be arbitrary, skewing the underlying structure of the data.

In order to obtain information about overall structures, multidimensional scaling (MDS) is a better technique. This is because results of MDS analysis "represent the semantic dimensions underlying a domain of concepts" (Branaghan, 1990, p. 111). However, as is the case with Pathfinder, MDS gives no explicit grouping information among tested items. Chen (2003) indicates that proximity patterns "must be judged carefully to identify the underlying structure. Proximity-based pattern recognition is not easy and sometimes can be misleading" (p. 155). When the research goal is to group objects based on the underlying characteristics participants possess, cluster analysis is the most effective technique (Hair, et al., 2006, p. 559). Accordingly, I argue for cluster analysis over MDS in the present research project for two specific reasons: one is the weakness of MDS in distributing lexical items in a sensible way for a multidimensional space (i.e., spatial map), and the second is the superiority of cluster analysis in examining sorting task results.

MDS and cluster analysis have different benefits in regard to cognitive domain related data. MDS is for spatial data, and cluster analysis is for hierarchical data. Thus, when the underlying structure is believed to be spatial, MDS should be applied. In cases when the structure is believed to be hierarchical, cluster analysis is the right choice. The problem is that MDS often fails to reveal any latent structures. This held true with Rapoport and Fillenbaum (1972), in which MDS did not distribute HAVE verbs in an "intuitively sensible way in the space" (p. 127). As they indicated, this was because there is basically a taxonomic structure characterised by inclusion (i.e., hierarchical) relations among HAVE verbs. This suggests that the data should have been analysed by another statistical method to reveal the underlying structure in a more sensible way. In fact, Rapoport and Fillenbaum also ran cluster analysis and achieved a satisfactory result.

More importantly, MDS is weak at producing a distinct representation of the data it analyses and merely presents "the perceived relative image of a set of objects" (Hair, et al., 2006, p. 632). Because of this relative image issue, MDS sometimes fails to reveal any underlying structures in a sensible, distinctive way as was the case with Rapoport and Fillenbaum (1972). Rapoport and Fillenbaum reported that the results of MDS analysis were difficult to interpret regarding the configuration of HAVE verbs in cognitive structure.

This weakness of MDS was pointed out by Routh (1994) as well: "it is very clear that the points [of quantifiers] cannot be projected on to an obvious global dimension corresponding to quantity" (p. 207). Cluster analysis run on the same set of data revealed the distinctive features of the relationship. Routh (1994) stated that "most quantifiers are well defined in terms of unique features, but they also combine into clusters that are fairly interpretable" (p. 209). As discussed earlier, Miller (1969) and Rapoport and Fillenbaum (1972) also succeeded in revealing the underlying cognitive structures they were interested in by means of cluster analysis. The mathematical principle of cluster analysis is analysing the data to put individuals or groups into clusters. This fulfils the purpose of analysing the data a sorting task collects. It is safe to conclude that the underlying structure of sorting task results is examined and represented succinctly by cluster analysis.

Third, it should be noted that the analyses that Miller (1969), Rapoport and Fillenbaum (1972), Preece (1976) and Routh (1994) conducted were all concerned with group data. In their studies, individual participant data were combined into group matrices and submitted to different types of multivariate data analysis. They included MDS (Rapoport and Fillenbaum, 1972; Routh, 1994), cluster analysis (Miller, 1969; Rapoport and Fillenbaum, 1972; Routh, 1994) and Waern's method of graphic analysis (Preece, 1976). The Pathfinder analysis that Sanchez (2004) conducted was also one type of multivariate group data analysis. In these studies, structural analyses were confined to group data. Individual differences were not matters of concern and thus were beyond the scope of the analyses.

These studies addressing the issues of L1 cognitive structures did not regard the absence of individual data analysis as being problematic. What mattered was not how individual participants were similar to or different from each other, but rather how participants as a sample of a population (i.e., culture) behaved and revealed underlying cognitive structures. In adopting multivariate data analysis in lexical organisation studies, one should recognise this inherent group-oriented nature. However, in addressing issues of L1 and L2 lexical organisation, the presence or absence of individual differences between participants is crucial. Haastrup and Henriksen (2000), Wilks and Meara (2002) and Meara and Schur (2002) all addressed this issue.

Given the discussion above, there are two options. One would be to be concerned only with group data and not with individual differences (i.e., focus on group data analysis). This is not a mistaken idea as long as the research questions posed are confined to revealing how L1 and L2 participants are different from each other as a group. This is what Miller (1969), Rapoport and Fillenbaum (1972), Preece (1976) and Routh (1994) did, resulting in interesting findings.

Another choice would be to analyse data both by descriptive statistics (e.g., *t*-test and ANOVA) and multivariate data analysis. The results might be complementary or contradictory. However, it would be worth it to conduct these analyses because by doing so, experiments using sorting tasks could address both group and individual differences in lexical organisation.

2.12 Conclusion

The review of L2 lexical organisation research in this chapter confirmed that there are three research questions worth addressing in this thesis, those being: (1) whether L2 lexical organisation is more varied than that of the L1, (2) whether L1 lexical organisation has more, smaller, non-connected clusters than does L2 lexical organisation, and (3) whether the two types of lexical organisation are differently arranged from one another. All of these questions were derived from reviewing the findings and predictions of recent word association test-oriented lexical organisation studies such as Meara and Schur (2002), Wilks and Meara (2002) and Wilks, Meara and Wolter (2005). However, this thesis will aim to examine and reveal facets of L1 and L2 cluster structures using an elicitation technique other than word association tests.

The literature review indicates that sorting tasks are the most promising type of elicitation task in addressing the research questions identified above. However, the sorting tasks that Haastrup and Henriksen (2000) developed, which have been the only ones used in L2 lexical organisation research, have reliability limitations for the present research purpose. Their tasks had more to do with addressing the lexical knowledge of participants than with organisation. By reviewing L1 lexical organisation studies in the latter half of this chapter, it was found that among the types of sorting tasks, free sorting tasks best meet the purpose of the current study. Results would straightforwardly reflect the underlying cluster structures of the mental lexicons of participants.

Cross-sectional research is preferable to longitudinal in addressing issues of L1 and L2 differences in lexical organisation. Considering the purpose of this thesis, a longitudinal approach is beyond the scope of this project. A comparison of statistical analyses of solid cognitive structure research (Miller, 1969; Rapoport & Fillenbaum, 1972; Preece, 1976; Routh, 1994) showed that some multivariate analysis needs to be done on sorting task data to reveal how lexical organisation is structured. The literature review revealed that cluster analysis meets this purpose. It was also found that multivariate data analysis, including cluster analysis, is concerned exclusively with group data and not with individual differences. In this regard, descriptive statistics-based analysis (e.g., *t*-test and ANOVA) is also appropriate.

In conclusion, this study will attempt to reveal the clustering behaviours of L1 and L2 participants and their underlying organisational structures by means of free sorting tasks. A series of experiments will be carried out for this purpose. It is hoped that by doing so our understanding of L1 and L2 lexical organisations will be further enhanced.

Chapter 3: Replication of Haastrup and Henriksen (2000)

3.1 Introduction

The purpose of this thesis is to examine the cluster structures of L1 and L2 lexical organisations and how they are different from each other. In the literature review in Chapter 2, I argued that for this purpose a sorting task is a valid data elicitation technique. Haastrup and Henriksen (2000) was the only full study that incorporated sorting tasks into their research. Therefore, at the start of this thesis project, it is worthwhile to replicate their research and investigate it from the perspective of whether their sorting tasks have the validity to be used in a series of experiments to be conducted in this thesis. The replication was done using Haastrup and Henriksen's Primary Sorting Task (where participants sorted adjectives of emotion and physical dimension into four predetermined categories), the Card-Sorting Task (where participants sorted adjectives of emotion into four predetermined categories) and the Situation Task (where participants selected one or more adjectives in a sentence). The replication was not done by a longitudinal but rather cross-sectional approach, and thus this is only a partial replication of the original study.

Haastrup and Henriksen (2000) has some methodological limitations. Their analysis was mainly done on the results of individual words, although their original aim was to reveal issues regarding L2 network building. Haastrup and Henriksen were particularly concerned with Aitchison's (1994) idea that knowing the meanings of a word involves the ability to relate it to other words and "a full understanding of the meaning of many words requires a knowledge of the words which are found with it or related to it" (p. 63). However, their study focused on analysing how well Danish school children participants developed their knowledge of synonyms and near-synonyms as compared to the norm of British teenagers of the same age. Their study failed to address results in light of L2 lexical organisation as a whole. Thus, the study's weaknesses seemed to lie more in the way it analysed the results than in the tasks themselves.

Another purpose of this replication is to confirm whether I can find evidence to support the findings that Haastrup and Henriksen (2000) reported. Thus I will address whether participants find the Card-Sorting task the most difficult (as Haastrup and Henriksen's implicational scale-based analysis suggested) and whether participants in this replication do better in sorting adjectives of emotion (e.g., AFRAID and HAPPY) than in others (e.g., SAD and ANGRY), as was the case with the participants in Haastrup and Henriksen's study. It should be noted that their results were difficult to generalise because findings for only two of the 17 participants were analysed in depth. However, it would be worthwhile to confirm whether some categories of adjectives are actually easier than others for a non-Danish L2 participant group as well. In this regard, as in Haastrup and Henriksen, my investigation includes word-based analysis. Then I will discuss if the findings can be related to broader aspects of L2 lexical organisation.

Given the aims of this replication above, the following three research questions will be addressed:

(a) Are there any consistent relationships between the results of Tasks 1 (the Primary Sorting Task), 2 (the Card-Sorting Task) and 3 (the Situation Task) in terms of the lexical organisation of participants? Specifically, is Task 2 the most difficult among the three tasks?
(b) Are some categories of adjectives of emotion more difficult than others for participants?
(c) Are there any word specific factors that affect the performance of participants?

3.2 Method

3.2.1 Participants

Participants in this study were 30 first year-students (aged 18 to 19) at Kumamoto University, Kyushu, Japan. All were engineering majors and had been studying English for six or seven years. They were identified as intermediate-level EFL learners according to the number of years of their previous English learning and from answers to a questionnaire (asking for their scores on the TOEFL, TOEIC and other English proficiency tests) filled in at the time of the experiment. Thus, the participants had longer English learning experience and were assumed to have higher English proficiency than the Danish school children participants in Haastrup and Henriksen (2000).

3.2.2 Procedure

As shown above, the three tasks (the Primary Sorting Task, the Card-Sorting Task and the Situation Task) that Haastrup and Henriksen (2000) developed were adopted in this replication, being named Task 1, Task 2 and Task 3. In Task 1, participants were asked to place 39 adjectives of emotion and physical dimension into one of four categories: HAPPY, AFRAID, WEIGHT and SIZE, or TEMPERATURE. In Task 2, they were directed to sort 30 adjectives into four categories of adjectives of emotion (AFRAID, ANGRY, HAPPY, or SAD). One modification was made to the task directions. In Haastrup and Henriksen (2000), participants were directed to "just leave" unsorted words (i.e., the words they did not know and could not place in any category), but in the present replication, participants were asked to place such words into the fifth category labelled "unknown words". In Task 3, participants were required to select one or more adjectives from a set of four adjectives that were near-synonyms to an underlined adjective in a sentence. Task 3 had 16 questions. For details of Tasks 1, 2 and 3, see Appendices 3.1, 3.2 and 3.3. Participants were given these tasks in the form of a handout starting with Task 1, followed by Task 2 and then Task 3. They were instructed to work on the tasks in this order, and not to do again any tasks they had already completed. After piloting the tasks to students of the same proficiency level from the same university, the experiment was carried out on the 30 participants described above, using part of an English lesson. It took the participants 35 minutes on average to complete the three tasks.

3.3 Results

Results are reported in the three sections corresponding to each of the research questions in order. In section 3.3.1, the complete sets of data for each of the three tasks are compared to each other. Thus, regarding aspects of the results for the three tasks, (a) total scores, (b) adjectives of emotion (HAPPY and AFRAID) scores, (c) scores of HAPPY adjectives, and (d) scores of AFRAID adjectives are each submitted to a repeated measures ANOVA (Analysis of Variance). Also, correlations between all the tasks are calculated. In section 3.3.2, the results of Task 2 (the Card-Sorting Task) are investigated by comparing the sorting task results of the four adjective categories (i.e., HAPPY, ANGRY, AFRAID and SAD). In the analysis, descriptive statistics for each category of the results are calculated and compared to one other. Lastly, in section 3.3.3, the performance of the Danish and Japanese participants will be compared by examining the number and percentage of participants who sorted adjectives correctly in Tasks 1 and 2. Analysis is also done on word specific factors that affected Task 2 results.

3.3.1 Between-task data analysis

To answer the first research question regarding the degree of relatedness between aspects of the three task results, four sets of statistical information will be shown. Each set is composed of tables of (a) means and standard deviations of task results, (b) repeated measures ANOVA results, and (c) multiple comparisons of the three task results.

First, the total scores of Tasks 1, 2 and 3 were compared. Each task had a different maximum score (Task 1 = 39; Task 2 = 30; Task 3 = 38). Thus, the comparison was made by means of converting the scores of the three tasks to percentage scores. The mean percentage scores and *SD*s are shown in Table 3.1.

Table 3.1. Mean percentage scores and standard deviations of task results: Total score

	Task 1 ($n = 30$)	Task 2 ($n = 30$)	Task 3($n = 30$)
Mean	68.72	47.33	48.42
SD	4.44	6.80	9.08

Table 3.2 summarises a repeated measures ANOVA result related to the results tabulated in Table 3.1.

Table 3.2. Repeated measures ANOVA results: Total score

Sources of variance	Sums of Square	*df*	Mean Square	*F*
Task type	8704.49	2	4352.24	87.99**
Error	4303.48	87	49.47	
Total	13007.97	89		

Note. ** $p < 0.01$.

Table 3.2 reveals that there was a statistically significant difference between the three tasks. To search for between-task differences, multiple comparisons were run using the Scheffé test. Table 3.3 contains the results of the Scheffé test run on the data in Table 3.1.

Table 3.3. Multiple comparisons of three task results (Scheffé test): Total score

Task pair compared	Mean difference	Critical value	*p*-value
Task 1 vs. Task 2	21.39**	4.43	< .0001
Task 1 vs. Task 3	20.30**	4.43	< .0001
Task 2 vs. Task 3	1.09	4.43	0.8361

Note. ** $p < 0.01$.

Table 3.3 shows that there were statistically significant differences between the Task 1 and Task 2 results and that Task 1 was easier for participants. Similarly, there was a significant difference between the Task 1 and Task 3 results and that Task 1 was easier than Task 3. There was no significant difference between the Task 2 and Task 3 results, suggesting that the difficulty of the two tasks was about the same for participants. Thus, the present replication results were different from Haastrup and Henriksen's (2000), which suggested that Task 2 (the Card-Sorting Task) was the most difficult among the three tasks.

Second, the total scores of the seven adjectives of emotion, which were used in all three tasks, were analysed. Table 3.4 displays the means and standard deviations for each of the three tasks.

Table 3.4. Means and standard deviations of task results: Seven adjectives of emotion

	Task 1 (n = 30)	Task 2 (n = 30)	Task 3(n = 30)
Mean	6.87	6.17	4.53
SD	0.35	0.95	1.57

Note. Maximum = 7.

Table 3.5 summarises a repeated measures ANOVA run on the results contained in Table 3.4.

Table 3.5. Repeated measures ANOVA results: Seven adjectives of emotion

Sources of variance	Sums of Square	*df*	Mean Square	*F*
Task type	86.02	2	43.01	37.01**
Error	101.10	87	1.16	
Total	187.12	89		

Note. $^{**}p < 0.01$.

A repeated measures ANOVA revealed that there was a statistically significant difference between the three tasks. To search for between-task differences, multiple comparisons were run using the Scheffé test. Table 3.6 contains the results of the Scheffé test run on the data in Table 3.4.

Table 3.6. Multiple comparisons of three task results (Scheffé test): Seven adjectives of emotion

Task pair compared	Mean difference	Critical value	p-value
Task 1 vs. Task 2	0.700*	0.693	0.0472
Task 1 vs. Task 3	2.333**	0.693	< .0001
Task 2 vs. Task 3	1.633**	0.693	< .0001

Note. * < 0.05; $^{**}p < 0.01$.

Table 3.6 shows that there were significant differences between the results of each task pair comparison and that the differences between all the paired results were statistically significant. To put it in another way, the difficulty of the three tasks in descending order of difficulty was as follows: Task 1, Task 2 and Task 3. Again, the results did not support Haastrup and Henriksen's (2000) finding that Task 2 (the Card-Sorting Task) was the most difficult.

Third, the scores of HAPPY adjectives, which were used in all the tasks, were analysed. Table 3.7 displays the means and standard deviations for each of the three tasks.

Table 3.7. Means and standard deviations of task results: HAPPY adjectives

	Task 1 ($n = 30$)	Task 2 ($n = 30$)	Task 3($n = 30$)
Mean	3.93	3.93	3.00
SD	0.25	0.25	0.95

Note. Maximum = 4.

Table 3.8 summarises a repeated measures ANOVA run on the results contained in Table 3.7.

Table 3.8. Repeated measures ANOVA results: HAPPY adjectives

Sources of variance	Sums of Square	*df*	Mean Square	*F*
Task type	17.42	2	8.71	25.49**
Error	29.73	87	0.34	
Total	47.15	89		

Note. $^{**}p < 0.01$.

As Table 3.8 shows, a repeated measures ANOVA revealed that there was a statistically significant difference between the three tasks. To search for between-task differences, multiple comparisons were run using the Scheffé test. Table 3.9 contains the results of the Scheffé test run on the data in Table 3.7.

Table 3.9. Multiple comparisons of three task results (Scheffé test): HAPPY adjectives

Task pair compared	Mean difference	Critical value	*p*-value
Task 1 vs. Task 2	0.000	0.376	0.0000
Task 1 vs. Task 3	0.933**	0.376	< .0001
Task 2 vs. Task 3	0.933**	0.376	< .0001

Note. $^{**}p < 0.01$.

Table 3.9 reveals that there was a statistically significant difference between Task 1 and Task 3 and Task 2 and 3. That is, Tasks 1 and 2 turned out to be easier than Task 3.There was no significant difference between the Task 1 and Task 2 results, suggesting that the difficulty of the two tasks was about the same for participants. Again, these results did not support the findings of Haastrup and Henriksen (2000), which implied that Task 2 (the Card-Sorting Task) was the most difficult.

Fourth, the scores of AFRAID adjectives, which were used in all three tasks, were analysed. Table 3.10 shows the means and standard deviations for each of the three tasks.

Table 3.10. Means and standard deviations of task results: AFRAID adjectives

	Task 1 ($n = 30$)	Task 2 ($n = 30$)	Task 3($n = 30$)
Mean	2.93	2.23	1.53
SD	0.25	0.94	0.90

Note. Maximum = 3.

Table 3.11 summarises a repeated measures ANOVA run on the results contained in Table 3.10.

Table 3.11. Repeated measures ANOVA results: AFRAID adjectives

Sources of variance	Sums of Square	*df*	Mean Square	*F*
Task type	29.40	2	14.70	25.23**
Error	50.70	87	0.583	
Total	80.10	89		

Note. $^{**}p < 0.01$.

A repeated measures ANOVA revealed that there was a statistically significant mean difference between the three tasks. To search for between-task differences, multiple comparisons were run using the Scheffé test. Table 3.12 contains the results of the Scheffé test run on the data in Table 3.10.

Table 3.12. Multiple comparisons of three task results (Scheffé test): AFRAID adjectives

Task pair compared	Mean difference	Critical value	*p*-value
Task 1 vs. Task 2	0.70*	0.491	0.0028
Task 1 vs. Task 3	1.40**	0.491	< .0001
Task 2 vs. Task 3	0.70*	0.491	0.0028

Note. $^{*}p < 0.05$; 1$^{**}p < 0.01$.

Table 3.12 shows that there were significant differences between the results of each task pair comparison and that the differences between all these results were statistically significant. That is, the difficulty of the three tasks in descending order of difficulty was as follows: Task 1, Task 2 and Task 3. Once again it was found that the present results were not consistent with Haastrup and Henriksen (2000), who reported that Task 2 was the most difficult among the three tasks.

The analyses thus far have revealed that there is a tendency of Task 1 being the easiest, followed by Task 2 and Task 3. This is in direct contradiction to the finding of Haastrup and

Henriksen (2000) that Task 2 (the Card-Sorting Task) was the most demanding and difficult among the three tasks. In sum, the present replication results do not support Haastrup and Henriksen (2000).

Lastly, Table 3.13 tabulates the correlations between all of the sets of data analysed above.

Table 3.13. Correlations between tasks

	1	2	3	4	5	6	7	8	9	10	11	12
1 Task 1 total	1.00											
2 Task 2 total	-0.25	1.00										
3 Task 3 total	-0.27	0.42 **	1.00									
4 Task 1happy	0.20	0.09	-0.05	1.00								
5 Task 2happy	0.28	-0.04	-0.13	0.46 **	1.00							
6 Task 3happy	0.02	0.41 *	0.61 **	-0.14	0.00	1.00						
7 Task 1afraid	0.28	-0.11	-0.40 *	-0.07	-0.07	-0.14	1.00					
8 Task 2afraid	0.31	0.30	-0.09	-0.22	-0.08	0.27	0.07	1.00				
9 Task 3afraid	0.07	0.20	0.47 **	-0.44 *	-0.14	0.45 *	0.16	0.30	1.00			
10 Task 1 happy+afraid	0.36	-0.01	-0.33	0.68 **	0.29	-0.21	0.68 **	-0.11	-0.21	1.00		
11 Task 2 happy+afraid	0.38 **	0.28	-0.13	-0.10	0.19	0.27	0.05	0.96 **	0.26	-0.03	1.00	
12 Task 3 happy+afraid	0.05	0.36 *	0.64 **	-0.34	-0.08	0.86 **	0.01	0.34	0.84 **	-0.25	0.31	1.00

Note. * $p < 0.05$; ** $p < 0.01$.

Table 3.13 shows that there was a moderate correlation between "Task 2 total" and "Task 3 total" ($r^2 = 0.42, p < 0.01$). However, for the other pairs of task totals, there was no significant correlation observed. It should be noted that the correlations between "Task 1 total" and "Task 2 total" and between "Task 1 total" and "Task 3 total" were not statistically significant and their correlation indices were negative ($r^2 = -0.25, p = 0.181$ for the former and $r^2 = -0.27$, $p = 0.152$ for the latter). Thus, these sets of correlation measures cannot be judged to be related to each other. The results did not support Haastrup and Henriksen's (2000) claim of an implicational relationship between tasks. For example, the present experiment failed to show that if a participant "could successfully carry out the Card-Sorting Task, then he/she could also carry out the Primary Sorting Task, but the reverse was not the case" (p. 233). The same tendency was observed between the other pairs of tasks as well. Particularly, this held true with the correlation indices between scores of the seven adjectives of emotion ("happy + afraid") that were used in all three tasks. None of the indices were significant and two of them were even negative. For example, the correlation between "Task 1 happy + afraid" and "Task 2 happy + afraid" was -0.03, $p = 0.856$ and the one between "Task 1 happy + afraid" and "Task 3 happy + afraid" was -0.25, $p = 0.193$. In sum, the tasks do not seem to be related to each other to the degree that they have an implicational relationship of the sort that Haastrup and Henriksen argued was important in network building.

3.3.2 Between-category data analysis

In this section, results related to the second research question (i.e., are some categories of adjectives of emotion more difficult for participants in completing tasks?) will be reported. As was the case with Haastrup and Henriksen (2000), analysis was conducted on the results of Task 2 (the Card-Sorting Task).

Table 3.14 shows the descriptive statistics of the sorting results of the four categories.

Table 3.14. Task 2 (Card-Sorting Task) results: Descriptive statistics

	AFRAID	ANGRY	HAPPY	SAD
Total items (*k*)	9	7	9	7
Mean (*M*)	4.60	0.80	5.93	2.87
M/*k* (%)	51.11	11.43	65.89	41.00
mode	5	0	6	3
Mdn	5	0	6	3
Range	8	4	2	5
SD	2.13	1.10	0.58	1.63

Note. Following the scoring keys of Haastrup and Henriksen (2000), *uneasy* sorted into both AFRAID and SAD as well as *moody* sorted into both ANGRY and SAD were identified to be correct.

Comparing the mean scores of participants who correctly sorted adjectives into their proper categories, it was found that the HAPPY category was the easiest to sort, followed by the AFRAID and the SAD categories. Meanwhile, the ANGRY category was extremely difficult for participants to sort into in the present replication. The mean score for this category was only 0.80 out of a possible maximum score of seven and the mode was zero. Eleven participants out of 30 got no answers correct. Thus, the results only partially support Haastrup and Henriksen (2000) since the researchers reported that the AFRAID category was as easy as the HAPPY category to sort.

3.3.3 Word-based data analysis

In this section, answers will be sought to the third research question of whether there are some word specific factors affecting the performance of participants. For that purpose, the number and percentage of participants who sorted adjectives correctly in Tasks 1 and 2 were calculated and the results are shown in Table 3.15. The table juxtaposes the results of both the present replication and Haastrup and Henriksen (2000) for comparison. The table reveals that the Japanese EFL learners performed far better than their Danish counterparts in every word in both tasks (except *scared* in Task 1). This is not surprising since the Japanese EFL learners (first year college students) had had a longer time learning English and should have developed higher English proficiency, acquired more English vocabulary and built up better lexical organisation of English adjectives in their mental lexicons.

Table 3.15. Number and percentage of participants sorting adjectives correctly in Tasks 1 and 2 (Danish, n = 17; Japanese, n = 30)

	Task 1				Task 2			
	DNS		JPN		DNS		JPN	
	No.	%	No.	%	No.	%	No.	%
glad	17	100.0	30	100.0	17	100.0	30	100.0
pleased	16	94.1	30	100.0	12	70.6	30	100.0
cheerful	12	70.6	28	93.3	10	58.8	29	96.7
excited	10	58.8	30	100.0	11	64.7	29	96.7
scared	17	100.0	29	96.7	10	58.8	23	76.7
frightened	15	88.2	30	100.0	10	58.8	22	73.3
anxious	10	58.8	29	96.7	6	35.3	22	73.3
Average	13.9	81.5	29.4	98.1	10.9	63.9	26.4	88.1

Note. DNS = Danish; JPN = Japanese.

Table 3.15 shows that Japanese EFL learners generally performed better in Task 1 than in Task 2, as was the case with Danish learners of English. This was particularly true for four adjectives (*excited, scared, frightened,* and *anxious*) out of the seven adjectives tested. Meanwhile, more participants sorted *cheerful* correctly in Task 2 than in Task 1. Regarding the two adjectives *glad* and *pleased,* all participants succeeded in correctly sorting them in both tasks. These two words were obviously well known to all participants, but had no sensitivity to distinguish participants' performance between the two tasks.

Second, in order to search for other word specific factors affecting sorting results, tested adjectives in Task 2 were ranked in increasing order of difficulty (Table 3.16) for Japanese participants. The table shows adjectives, categories, and the number and percentage of participants who sorted the words correctly. In the table, word levels in the Word List of Hokkaido University (Hokudai Word List; Hokudai) indicating which school level of English learning in Japan each adjective belongs to are shown. BNC word frequency (word frequency per million words in the British National Corpus) is juxtaposed with it. They are used to investigate whether word level and frequency play a role in participants' sorting performance.

Table 3.16. Task 2 Results: Adjectives ranked in order of difficulty (Japanese participants, n = 30)

No.	*Adjective*	CATEGORY	No. of correct participants	% of correct participants	Hokudai	BNC
1	*glad*	HAPPY	30	100	[1]	41
1	*pleased*	HAPPY	30	100	[1]	47
3	*cheerful*	HAPPY	29	96.7	[2]	12
3	*excited*	HAPPY	29	96.7	[5]	16
5	*high*	HAPPY	28	93.3	[1]	574
6	*overjoyed*	HAPPY	26	86.7	[5]	**
7	*scared*	AFRAID	23	76.7	[5]	12
8	*anxious*	AFRAID	22	73.3	[2]	31
8	*frightened*	AFRAID	22	73.3	[5]	6
10	*miserable*	SAD	21	70.0	[3]	12
11	*terrified*	AFRAID	19	63.3	[5]	**
11	*disappointed*	SAD	19	63.3	[5]	21
13	*uneasy*	AFRAID	18	60.0	[3]	**
14	*sorrowful*	SAD	17	56.7	[3]	**
15	*panic-stricken*	AFRAID	16	53.3	[5]	**
16	*depressed*	SAD	13	43.3	[5]	15
17	*alarmed*	AFRAID	11	36.7	[5]	**
18	*annoyed*	ANGRY	10	33.3	[5]	**
19	*moody*	SAD	9	30.0	[4]	**
20	*distressed*	AFRAID	6	20.0	[5]	**
21	*mad*	ANGRY	5	16.7	[2]	32
21	*thrilled*	HAPPY	5	16.7	[5]	**
23	*upset*	SAD	4	13.3	[2]	18
24	*outraged*	ANGRY	3	10.0	[5]	**
24	*uneasy*	SAD	3	10.0	[3]	**
26	*cross*	ANGRY	2	6.7	[5]	**
26	*grumpy*	ANGRY	2	6.7	[2]	**
28	*petrified*	AFRAID	1	3.3	[5]	**
29	*furious*	ANGRY	1	3.3	[3]	13
30	*moody*	ANGRY	1	3.3	[4]	**
31	*chuffed*	HAPPY	1	3.3	[5]	**
32	*elated*	HAPPY	0	0.0	[5]	**

Note. Hokudai = Word level in the Word List of Hokkaido University; [1] = junior high school level; [2] = senior high school level; [3] = senior high school graduate level; [4] = university basic level; [5] = word level lower than level 4; BNC = British National Corpus frequency (i.e., frequency per million words in the BNC). The symbol ** indicates that the frequency of the adjective in question is less than one per million words in the BNC.

Taking a look at the results, the top six adjectives in Table 3.16 (all of which more than 26 (86.7%) participants sorted correctly) all fell into the HAPPY category: *glad, pleased, cheerful, excited, high* and *overjoyed*. Furthermore, five of them are high frequency words, including *excited* (BNC = 16), whose verb form *excite* is a Level 2 word in the Hokudai Word List. In addition, the 6th-ranked adjective *overjoyed* is not a high frequency word, but the stem *joy* is a high frequency word with a BNC index of 27 and is a senior high school level word according to the Hokudai Word List. This seems to account for why 26 participants were able to sort *overjoyed* correctly. *Excite* is a high frequency, senior high school level word in Japan. *Scared* (ranked 7th), *anxious* (ranked 8th) and *frightened* (also ranked 8th), all part of the AFRAID category, are also all high frequency words. Stem forms of *scare* for *scared* and *frighten* for *frightened* are both high frequency, level 2 words. Thus the results revealed that the word frequency of an adjective (as well as its stem form) and when a word is learned in the Japanese school system were both crucial factors in how well participants were able to sort the HAPPY and AFRAID adjectives. On the other hand, Table 3.16 reveals that there are also adjectives in HAPPY and ARRAID that many participants failed to sort correctly and they were lower frequency, rare words (e.g., *chuffed* and *elated* in HAPPY and *distressed* and *petrified* in AFRAID). More importantly, adjectives in the SAD and ANGRY categories that many participants failed to sort correctly were also lower frequency, unfamiliar words (e.g., *depressed* and *upset* in SAD and *grumpy* and *cross* in ANGRY).

Taking all these factors into account, whether an adjective of emotion was sorted correctly in the present replication could be attributed to its word frequency and when it was learned. That is, there was no explicit word type effect on the sorting performance of participants (e.g., adjectives in HAPPY are easier to sort than those in ANGRY for participants).

3.4 Discussion

In this section, two points will be discussed. They are (a) variables contributing to L2 lexical knowledge development and (b) validity of Haastrup and Henriksen's (2000) sorting tasks for eliciting data in L2 lexical organisation research. The first point considers whether word based studies can be related to lexical organisation research. The second point involves a critical evaluation of the tasks Haastrup and Henriksen developed.

3.4.1 Variables contributing to L2 lexical knowledge development

First, word based findings in the present replication will be discussed as Haastrup and Henriksen did. This is because it would be meaningful to confirm whether the present replication supports Haastrup and Henriksen's findings in word based analysis. Then, it will be discussed whether word based analysis can shed light on lexical organisation issues. Adjectives in the HAPPY and AFRAID categories (seven in total) that were used in all three tasks will be focused on.

The sorting results of the adjectives in the HAPPY category (i.e., *glad, pleased, cheerful, excited*) were all better than those of the adjectives in the AFRAID category (i.e., *scared, frightened, anxious*) in Task 2 (Card-Sorting Task). This supports the results of Haastrup and Henriksen (2000). What factors brought about this seemingly clear-cut result? To answer this question, I focused on how frequent the seven adjectives are and at which school level Japanese EFL learners learn them (Table 3.17 below with data taken from Table 3.16). As predicted, it was revealed that how frequent a word is and how early participants learn it is a decisive factor in being able to sort the word correctly. Adjectives in the HAPPY category were higher frequency words than those in the AFRAID category (except *anxious*), and participants likely learned them at earlier stages of their English learning. This supports the analysis made regarding the increasing order of difficulty of tested words in the previous section. While the easiness of learning a word is affected by many variables, previous studies report that word frequency often plays a significant role in L2 lexical acquisition (e.g., Aizawa, 2003b; Brown, 1993; Nation 2001). The present results support this claim. The limitation is that this present analysis is limited to seven adjectives. This is a shortcoming in making word based analysis. Any claim based on such a small number of words cannot easily be generalised. Perhaps it would be fair to say that words selected for the HAPPY category just happened to be high frequency ones. Thus, the simple fact that some words in the task are from the HAPPY category does not lead to a conclusion that they are easier than others.

Table 3.17. Number and percentage of participants sorting HAPPY and AFRAID adjectives correctly in Task 2 (Japanese participants, $n = 30$)

No.	*Adjective*	CATEGORY	No. of correct participants	% of correct participants	Hokudai	BNC
1	*glad*	HAPPY	30	100.0	[1]	41
2	*pleased*	HAPPY	30	100.0	[1]	47
3	*cheerful*	HAPPY	29	96.7	[2]	12
4	*excited*	HAPPY	29	96.7	[5]	16
5	*scared*	AFRAID	23	76.7	[5]	12
6	*frightened*	AFRAID	22	73.3	[5]	6
7	*anxious*	AFRAID	22	73.3	[2]	31

Note. Hokudai = Word level in the Word List of Hokkaido University; [1] = junior high school level; [2] = senior high school level; [3] = senior high school graduate level; [4] = university basic level; [5] = word whose level is lower than level 4; BNC = British National Corpus frequency (i.e., frequency per million words in the BNC).

On the other hand, one of the AFRAID adjectives turned out to be the most difficult for participants to sort correctly for another reason. It is concerned with polysemous meanings of the word *anxious* which is not a low frequency word (BNC = 31). It has the two main meanings "eager" and "worried (afraid)". The two meanings seemed to be so competitive with each other in the mental lexicons of the present Japanese participants that many of them could not sort the word correctly. Some studies argue for the difficulty of learning words that have polysemous meanings (e.g., Bensoussan & Laufer, 1984; Furuie, 2003; Laufer, 1997), and the present replication provides evidence for these studies. However, it should be noted that this is only a single piece of evidence and has only a limited impact on the overall sorting performance analysed earlier. In addition, many high frequency words are essentially polysemous. Moreover, it is very difficult to analyse the polysemous nature of words one by one and to discuss how decisive an impact this lexical feature has on overall sorting performance and aspects of the lexical organisation. This difficulty will be multiplied when we extend the scope of analysis to other word specific factors. They include L1 and L2 cognates, word stem and affix information, and the effects of input from such sources as TV, movies and the Internet on individual word acquisition. Certainly we could make a long list of factors involving lexical knowledge for individual words, although it would be hard work. But it is doubtful whether we could draw a wider picture of lexical organisation through such analysis, no matter how many pieces of detailed information we could collect. Results gained by such analysis are difficult to relate to facets of lexical organisation. I argue that word based analysis can shed little light on lexical organisation issues.

Besides the individual word analysis shown above, another important point of interest is whether the sorting results of HAPPY and AFRAID adjectives as semantic fields had a significant impact on the overall task results and the degree of participants' network building. To make the analysis simpler, attention should be drawn again to the correlation indices between each pair of task results summarised in Table 3.13. If some significant correlation is found (e.g., between "Task 1 afraid" and "Task 2 total"), it would suggest that the sorting results of a group of three AFRAID adjectives can be related to L2 lexical development. Through such an investigation, I found four statistically significant moderate correlations among 18 potential pairs: "Task 3 happy" and "Task 2 total" ($r^2 = 0.41, p < 0.05$), "Task 3 happy" and "Task 3 total" ($r^2 = 0.61, p < 0.01$), "Task 1 afraid" and "Task 3 total" ($r^2 = 0.40, p < 0.05$) and "Task 3 afraid" and "Task 3 total" ($r^2 = 0.47, p < 0.01$). Two of these results were part-whole correlations and thus they should be treated with care (because part-whole correlations are usually higher than we would expect). However, these results contained an interesting feature. It was revealed that the results of the HAPPY adjectives in Task 3 (the multiple-choice test to measure synonym knowledge) had a moderately significant correlation with those of Task 2 (the Card-Sorting task). If this finding could be generalised, the development of the lexical organisation of HAPPY adjectives could be considered to be an index of L2 lexical knowledge. Of course, this claim is too optimistic to be convincing. The HAPPY adjectives used in the sorting task accounted for only 13.3% (4 words out of a total of 30 words), and they were all high frequency words that 29 or all 30 participants correctly sorted. Thus the number of HAPPY adjectives was too small to have the sensitivity to tap into lexical organisation issues. This is an inherent fundamental weakness in making word based analysis. The results of the present replication were more a matter of "chance" than anything else. This held true with the other three significant correlations as well. That is, there were too few tested words for us to be able to claim that the results can be generalised. All in all, it should be noted that statistically significant correlations account for only 22.2% in the correlation matrix (4 pairs out of 18 potential pairs).

In sum, the analysis above confirmed that it is extremely difficult to detect features to be generalised in word based analysis and then be able to relate them to an examination of lexical organisation. This suggests that if we continue to pursue word based studies, little will be revealed regarding structures of lexical organisation.

3.4.2 Validity of Haastrup and Henriksen's (2000) sorting tasks for lexical organisation studies

In this section, I will address the issue of whether the sorting tasks used in this replication have validity for conducting lexical organisation studies and in particular the series of experiments in the present thesis project. Thus, Task 1 (the Primary Sorting Task) and Task 2 (the Card-Sorting Task) will be discussed in terms of how valid they are in empirically testing the idea that "vocabulary acquisition is more a matter of system learning than of item learning" (Haastrup and Henriksen, 2000, p. 225).

Task 1 (the Primary Sorting Task) was designed to tap into the L2 lexical structures of participants, as was the case with Task 2 (the Card-Sorting Task). The task attempted in particular to reveal how well semantic fields (i.e., adjectives of emotion and physical dimension) are integrated in L2 lexical organisation. Accordingly, participants were directed to sort adjectives into the four categories of WEIGHT and SIZE, TEMPERATURE, HAPPINESS and ANXIETY. The purpose of the study was fine, but the task lacked in content validity since the lexical areas covered in it could not address "the representativeness of our measurement regarding the phenomenon about which we want information" (Mackey & Gass, 2005, p. 107). This is related to the question: "Does your experiment really measure what you intended?" (Wray, Trott, & Bloomer, 1998, p. 163) when one designs an experiment of this sort. With regard to adjectives of physical dimension, the task addressed only two categories among many dimensions. Why were they chosen? To put it in another way, why cannot speed, force, distance, power, etc. be the representative of the physical dimension? The study did not address this concern. Thus the scope of the research became very narrow. Eventually, as is often the case with many lexical acquisition studies, the findings could be applied "only to very specific sets of words" (Meara, 2002, p. 403). That is, there appeared to be no explanation of why WEIGHT and SIZE along with TEMPERATURE were chosen as representative categories for the experiment. Compared with adjectives of the physical dimension, adjectives of emotion might have a shorter list of categories (e.g., the semantic fields of happiness, pleasantness, sadness, anxiety, etc.). Therefore, using the two categories of HAPPY and AFRAID in the Primary Sorting Task might have been less problematic than using those of the physical dimension. However, even if that logic held true, the study should have stated the reason why these two categories were selected for Task 1. Unfortunately, there was no mention of this essential issue in the paper, and it thus reduces the value of the study.

Another shortcoming is that tested words in the four categories were heavily skewed in number. The task had 17 words in the WEIGHT and SIZE category, 8 in the TEMPERATURE category, 7 in the HAPPINESS category and 8 in the ANXIETY category. Thus, the task was composed of 25 adjectives of the physical dimension and 15 words of adjectives of emotion. This imbalance between the categories tested made it difficult to compare and analyse the results. The number of adjectives in each of the four categories should have been equal to each other to run reliable analyses.

One more weakness of Task 1 is the lack of a balanced distribution of words between categories in terms of word frequency. Haastrup and Henriksen (2000) did not give explicit information concerning the word list against which the tested adjectives were checked for word frequencies. As far as I could tell, the study did not seem to have adopted any particular word list as a criterion in choosing the words tested. Therefore, I adopted the JACET 8000 (Japan Association of College English Teachers List of 8000 Basic Words) to check which level each adjective used in the task was at. Table 3.18 tabulates the frequency information of the 40 adjectives used in the study.

Table 3.18. Answer keys to Task 1 (Primary Sorting Task) and word frequencies in the JACET 8000 (k = 40)

Type	CATEGORY	Level								
		1	2	3	4	5	6	7	8	Lower
Adjectives of the physical dimension	WEIGHT and SIZE (17 words)	*skinny*	*fat*	*spare*		*slim*	*slender*	*plump*		*anorexic*
			lean					*stout*		*chubby*
			thin							*flabby*
										podgy
										portly
										scrawny
										tubby
										wiry
	TEMPERATURE (8 words)	*cold*			*chilled*					*lukewarm*
		hot			*freezing*					*scalding*
		warm								*tepid*
Adjectives of emotion	HAPPINESS (7 words)	*happy*	*glad*	*cheerful*			*jolly*			
			pleased	*excited*						
				gay						
	ANXIETY (8 words)	*afraid*	*nervous*	*anxious*		*uneasy*				
			worried	*excited*						
				frightened						
				scared						

Note. JACET 8000 = Japan Association of College English Teachers List of 8000 Basic Words. Words at the "Lower" level are those that are not found on the list of JACET 8000. In identifying word levels, a stem form was used in the case that the adjective form in question was not found in the JACET 8000 (e.g., *skin* for *skinny*). Credit for the word *excited* was given if it was sorted into the HAPPINESS or ANXIETY category as Haastrup and Henriksen (2000) did. Answer keys are based on the responses British teenagers gave.

As the table shows, the distribution of the tested words regarding word frequency (i.e., level) in Task 1 was extremely skewed, and the word ratios at each level were very different from one category to another. Most noticeably, 11 of the 25 (44.0%) adjectives in the physical dimension were at the "Lower" level, meaning that they are not on the JACET List of 8000 Basic Words. Meanwhile, none of the 15 adjectives of emotion fell into the "Lower" level. Thus, the adjectives of emotion were composed of far more high frequency words than were the adjectives of the physical dimension. Thus, the sorting task results of the adjectives in the four categories in Task 1 cannot be reliably compared to each other, and an analysis from any perspective seems to be invalid.

Haastrup and Henriksen (2000) reported that "the Primary Sorting Task [Task 1] was simple to score, in that there was full agreement within the reference group" (p. 231). This suggests that even such words as *anorexic, scrawny* and *wiry*, which most intermediate-level EFL learners rarely encounter, were well known to all of the British teenager participants. Of course, this does not mean that those rare words have valid "representativeness of measurement" addressing L2 lexical organisation issues. Actually, the Japanese college students in the present replication sorted them correctly into their correct categories at a very low rate; 4 (13.3%) for *anorexic*, 1 (3.3%) for *scrawny* and 4 (13.3%) for *wiry*. I argue that words tested in psycholinguistic research studies should be ones participants know the meanings of. Otherwise, a task will end up being a test of lexical knowledge rather than a test that taps into underlying lexical organisation. Unfortunately, Task 1 in Haastrup and Henriksen (2000) and the present replication turned out to be a lexical knowledge test that included a significant number of rare words that were beyond the level of the basic word list of the JACET 8000. This implies that it is requisite for researchers to use words whose meanings are known to participants in designing sorting tasks to address the lexical organisation (i.e., network building) of the mental lexicons of the participants.

Taking all these points into consideration, I contend that Task 1 (the Primary Sorting Task) has several problems and is not suitable for lexical organisation studies.

Task 2, the Card-Sorting Task, was a key data elicitation technique in addressing the network building issues in Haastrup and Henriksen (2000). The task required participants to sort 30 adjectives of emotion into four categories, and the results were assumed to reflect their "interlanguage network in that it depends solely on relational aspects of meaning" (p. 229). What was attempted in Task 2 was worthwhile, but the task also had some limitations as Task 1 did. They were concerned with task validity.

First, Task 2 gave participants an unnecessary explanation that led them to activate their conceptual knowledge more than their L2 lexical knowledge. Participants were given the explicit explanation that tested words were all adjectives that describe emotions. Then they were asked to sort them into four categories while working out category names for themselves. Thus, participants could easily predict that the 30 words they were sorting were words of happiness, anger, sadness, anxiety, etc. It should be noted that the explanation was more related to the conceptual knowledge people have developed through cognitive development and L1 vocabulary acquisition. Such knowledge is assumed to be shared with L2 lexical knowledge once the latter has been acquired. However, if L2 lexical knowledge has not been satisfactorily acquired yet, telling participants that tested words are adjectives of emotion gives them unnecessary assistance and therefore lessens the task reliability.

As discussed earlier, the categories for adjectives of emotion are limited in number, and it would be far easier for participants to predict what categories are being tested when they are given an explicit task explanation like the one above. Thus, the task primarily tapped into the conceptual knowledge of participants. Therefore, I argue that Task 2 lacked in validity as a task to probe into L2 lexical organisation. To increase the validity of Task 2, it would have been desirable to simply have explained to participants that the words tested were all adjectives and then have directed them to sort the words without further explanation. With such direction, participants would have been required to work out which lexical categories tested words were a part of, thus activating their underlying L2 lexical knowledge and organisation that they had already developed. The results would have produced more reliable data regarding network building of adjectives of emotion.

Second, and related to the first limitation, fixing the number of sorting categories in advance has more disadvantages than advantages when a sorting task poses questions of lexical organisation. There is of course one particular merit that fixed sorting has. Restricting the number of categories that participants are allowed to make "has the advantage of standardizing the variance of the sorting categories" (Coxon, 1999, p. 20). Therefore, fixed sorting is a suitable technique in studying participants' subjectivity in such areas as political science and health science. Meanwhile, lexical organisation research primarily attempts to address issues such as whether there is variability in the task results among participants. By using fixed sorting tasks, one cannot investigate if such a variance between participants exists and how significant it is. Thus, I argue that a sorting task for lexical organisation research should not fix the number of sorting categories in advance. It would be more interesting for us to tap into how many clusters of words participants make and how large each one is. For that reason, free sorting outweighs fixed sorting.

Third, Task 2 lacked in content validity in that the number of tested words for each category was small and the distribution of the word frequencies of the words tested were too skewed to be able to produce reliable results. Table 3.19 shows the numbers and percentages of tested adjectives of emotion and their frequency level as identified by the JACET 8000.

Table 3. 19. Number and percentage of tested adjectives of emotion at each level of the JACET 8000

Type	CATEGORY	Level			Total
		1 to 4	5 to 8	Lower	
Adjectives of emotion	AFRAID	6	3	1	10
		60%	30%	10%	100%
	ANGRY	2	3	2	7
		29%	43%	29%	100%
	HAPPY	7	0	1	8
		88%	0%	13%	100%
	SAD	4	3	0	7
		57%	43%	0%	100%

Note. JACET 8000 = Japan Association of College English Teachers List of 8000 Basic Words. Words at the "Lower" level are those that are not found on the list of the JACET 8000.

While differences were not statistically significant (χ^2 (6) = 7.905, $p > 0.10$), the table shows that the ratios of words at a particular frequency level were rather different from one category to another. For example, regarding the highest frequency category Levels 1 to 4, AFRAID had 6 words, ANGRY had 2, HAPPY had 7 and SAD had 4. Meanwhile, at Levels 5 to 8, HAPPY had no word, whereas the other categories had three words each. Under this skewed distribution of the word frequencies of the tested words, the results could not be reliably analysed and could not offer evidence for generalisation. This shortcoming could also be found with rare words that are not a part of the basic word list of the JACET 8000 (Table 3.20).

Table 3.20. Answer keys to Task 2 (Card-Sorting Task) and the word frequencies in the JACET 8000 (k = 30)

Type	CATEGORY	Level								
		1	2	3	4	5	6	7	8	Lower
	AFRAID			*alarmed*		*distressed*				*petrified*
	(10 words)			*anxious*		*terrified*				
				excited		*uneasy*				
				frightened						
				panic-stricken						
				scared						
Adjectives of emotion	ANGRY		*mad*			*annoyed*				*cross*
	(7 words)		*moody*			*furious*				*grumpy*
						outraged				
	HAPPY	*high*	*glad*	*cheerful*						*chuffed*
	(9 words)		*overjoyed*	*excited*						*elated*
			pleased	*thrilled*						
	SAD		*moody*	*miserable*	*disappointed*	*depressed*		*sorrowful*		
	(7 words)		*upset*			*uneasy*				

Note. JACET 8000 = Japan Association of College English Teachers List of 8000 Basic Words. Words at the "Lower" level are those that are not found on the list of the JACET 8000. In identifying word levels, a stem form was used in case the adjective form in question was not found in the JACET 8000 (e.g., *mood* for *moody*). Following the scoring keys of Haastrup and Henriksen (2000), credit for the word *excited* was given if it was sorted into the HAPPY or AFRAID category. Similarly, credit was given if *uneasy* was classified into either AFRAID or SAD and if *moody* was classified into either ANGRY or SAD.

Table 3.20 shows that the AFRAID category contained *petrified*, ANGRY contained *cross* and *grumpy* and HAPPY contained *chuffed*. They were all adjectives that EFL learners were unlikely to have encountered. Perhaps it might have been less problematic if the data elicitation task had had a larger number of tested words for each category to make the results more reliable. In such a case, a few words unknown to participants would not have had much of an impact on their overall performance. However, when the total number of words in a task is small, each word affects the results significantly. Task 2 lacked in validity, especially in the choice of tested words.

Given the discussion of Tasks 1 and 2 above, it appears that both of the sorting tasks Haastrup and Henriksen (2000) developed were not valid enough to address issues of L1 and L2 lexical organisation. Both tasks had methodological shortcomings which weakened their ability to tap into the L2 network building issues the study attempted to explore. The present replication confirmed these limitations and highlighted the points which need to be improved for a valid sorting task to be used in experiments for lexical organisation studies.

3.5 Conclusion

The sorting tasks Haastrup and Henriksen (2000) developed both had weaknesses as data elicitation techniques for lexical organisation research. Needless to say, in so far as L2 lexical acquisition studies tend to place more focus on individual word knowledge development (e.g., synonyms and near-synonyms as in Haastrup and Henriksen, 2000), the tasks did play an important role. However, this is not the case with the present thesis project which attempts to explore issues of L1 and L2 lexical organisation.

As discussed in the literature review (Chapter 2), sorting tasks themselves are reliable tasks which are widely used in many disciplines. Haastrup and Henriksen's (2000) sorting tasks, especially their key Card-Sorting Task, can be useful in addressing issues of lexical organisation. In this regard, words used in a sorting task should be selected against some consistent criterion (e.g., a word list). Any kind of task restriction, such as fixing the number of sorting categories or telling participants the categories into which words can be sorted into in advance, should not be done.

In Chapter 4, a sorting task will be designed, taking the findings of this replication into consideration. Then the cluster structures of L1 and L2 lexical organisations and how they are different from each other will be explored further.

Chapter 4: Cluster number, size and variability

4.1 Introduction

In Chapter 3, I reported a replication of Haastrup and Henriksen (2000), the only full study that employed sorting tasks in L2 lexical organisation research. The replication revealed that their sorting tasks had a number of limitations as data elicitation techniques to tap into the cluster structures of L1 and L2 lexical organisations. The limitations were fourfold. First, the sorting tasks were concerned more with the lexical knowledge of the Danish school children, while containing many low frequency words they did not know. Second, the main sorting task (the Card-Sorting Task) tapped more into the conceptual knowledge of participants than into their lexical organisation. Third, their tasks were fixed sorting tasks that were less valid than free sorting tasks in lexical organisation studies. Fourth, in their sorting tasks, native speakers of English (British teenagers of the same age as the Danish participants) took part in the experiment as informants to provide correct answer keys for the task. Thus, related to the first problem above, the sorting results of L2 participants were examined by means of whether they were right or wrong against the L1 norm. Eventually, Haastrup and Henriksen's sorting tasks addressed how well L2 participants knew the tested words rather than how they developed L2 lexical organisation.

Taking all these limitations into account, I will report the first of five total experiments in this thesis. Tested words in the experiment were chosen from the ones that participants know well. I implemented a free sorting task where participants sorted the tested words into groups of words that they thought go together according to meaning. There were no restrictions regarding the number of clusters and words per cluster they could make. In addition, L1 English speakers participated in the experiment not as informants for correct answer keys in the task but as participants whose sorting results were compared to those of L2 participants' regarding aspects of lexical organisation.

In this chapter, I address three research questions regarding the differences in sorting behaviour between L1 and L2 participants: time to complete the sorting task, cluster number and size, and the variability. I include the issue of task completion time because it is an important variable to tell apart L1 participants from their L2 counterparts regarding their sorting behaviour (research question 1). Among dimensions of lexical competence, task completion time is related to the issue of accessibility which is mainly studied by means of lexical decision tasks in psycholinguistic studies (e.g., de Groot, 1993; Dijkstra, 2005; Francis, 2005; Ikemura & Kadota, 2003; Kroll, 1993; Kroll & de Groot, 1997; Nakanishi, 2003a; Sánchez-Casas & García-Albea, 2005; Smith, 1997). Haastrup and Henriksen (2000) did not pose a question of this kind in their sorting task based experiment. Therefore, we have had no information so far regarding which participant group (L1 or L2) takes less time to complete a sorting task while making a more rapid access to the mental lexicon and process more quickly the lexical information in sorting. This experiment is the first one that addresses the L1 and L2 differences of task completion time regarding participants' sorting behaviour.

Furthermore, I have another reason why the issue of task completion time should be included as a research question in the experiment. That is the discrepancy of task completion time of the participants between Meara and Schur (2002) and Wilks and Meara (2002). Both studies gave L1 and L2 participants a restricted word association test. In Meara and Schur (2002), L1 participants took 10 to 15 minutes to complete the test, whereas L2 participants "needed a full class period to complete the task" (p. 170). Meanwhile, in Wilks and Meara (2002), L1 participants took a longer time (an average of 11.25 min.) than their L2 participants (an average of 10.38 min.) to complete the task. Of course, the restricted WATs that both studies used did not have the same task directions and tested words. These different parameters might have led participants to make different levels of cognitive effort and take different lengths of time to complete the tests. However, as is the case with sorting tasks, they were both data elicitation tasks that tapped into the lexical organisations of the participants. Accordingly, the two tasks should have produced consistently similar tendencies in that either L1 or L2 participants took less time in both studies, but they didn't. Then how about the present sorting task experiment? Which of the two WAT experiments is similar to this experiment in task completion time? The answer will offer a clue to how much cognitive effort and time these psycholinguistic data elicitation tasks require participants to make. It is rare for experiments of L2 lexical organisation research to investigate task completion time. This is another reason why I decided to include task completion time into my research questions.

In this sorting task experiment, whether the number of clusters that L1 participants make is larger and the number of words per cluster that they make is smaller than their L2 counterparts is also addressed (research question 2). This is related to the finding of Meara and Schur (2002) that L1 lexical organisation had more, smaller, non-connected components (clusters) than did L2 lexical organisation. They interpreted the results as the reflection that L1 speakers are more aware of the semantic relations between words and how the words may fall into distinct sets. Their restricted WAT was sensitive in detecting the degree of participants' awareness of links among possible pairs of tested words. Will the same results be produced by a sorting task which asks participants to make clusters of words being related in meaning? If they do, the results can be generalised as an L1 and L2 difference in lexical organisation.

The third research question addresses whether L1 lexical organisation is less varied than L2 lexical organisation (research question 3). This L1 and L2 variability difference is what Meara and Schur (2002) confirmed in their experiment, giving firm support for previous studies that had reported this organisational dissimilarity (e.g., Meara, 1983; Postman & Keppel, 1970; Riegel & Zivian, 1972; Szalay & Deese, 1978). Does a sorting task find evidence for it as well? Can the L1 and L2 variability difference in lexical organisation be generalised? In a series of experiments using sorting tasks, I will attempt to answer these questions as well.

Given the aims of the first experiment in this thesis project, the following three research questions will be addressed:

1. Which participant group takes less time to complete a sorting task, L1 or L2 participants?
2. Do L1 participants make more clusters and fewer words per cluster than their L2 counterparts?
3. Is L1 lexical organisation less varied than its L2 counterpart?

4.2 Method

4.2.1 Participants

In this study, there were two participant groups. The first was comprised of 28 adult, native speakers of English (NS). They were teachers of English in Kumamoto, visiting scholars and students studying at Kumamoto University. The second consisted of 28 adult, advanced-level Japanese speakers of English (NNS). They were either English teachers at the college level or persons having a high competence of English as judged by a TOEFL score of 213 or more on the computer-based version or a score of 550 or more on the paper-based version or a TOEIC score of 730 or more that had been taken within the last two years.

4.2.2 Data collection

For the present experiment, I randomly selected 50 words from among the most frequent 500 words in the JACET 8000 (Japan Association of College English Teachers List of 8000 Basic Words): *ago, air, all, already, also, area, arm, arrive, believe, boy, business, century, clear, close, country, cry, doctor, dream, early, figure, find, form, help, history, keep, law, lot, matter, nature, next, nothing, open, our, person, place, police, power, president, reason, shop, social, stand, step, street, then, understand, very, walk, war, while*. When identified as the word categorised in the first word class in the word list, they consisted of 30 nouns (60%), five verbs (10%), six adjectives (12%), six adverbs (12%), one pronoun (*nothing*) (2%), one determiner (*our*) (2%) and one conjunction (*while*) (2%).

After piloting, the task named Card-sorting Game was administered to the participants individually. Each pile of 50 cards, on which the individual English words were printed, was shuffled and bound with a rubber band and put into an envelope for each participant. Each card was 2.2 by 4.4 centimetres. Participants were invited to sort them into groups of words that they thought would go together according to meaning. There were no rules for the task except that participants could not sort a word by its word class. They might have found words that they thought didn't seem to fit into any of the word groups they made. They could leave these as single cards or individual groups of single, "isolated" words. It didn't matter how many groups they made. I set a 20-minute time limit to complete the task, but participants could finish the task earlier if they were happy with the results. The directions were printed on a piece of A4-size paper (see Appendix 4.1). They were written in English for native speakers of English, and in Japanese for Japanese participants. Participants read them through and asked questions, if any, and then worked on the task. They got a small present when they finished.

The time that participants needed to read the cards and to sort them were recorded all together, and the total was regarded as the time they took to complete the sorting task. This was timed from when they started to read the cards to when they stopped sorting the cards and announced that they had finished. However, the distinction between reading and sorting the cards was not recorded. The reason for doing so is as follows. Observing participants in the pilot test, it was not always clear-cut when a participant started to sort the cards. Some of them sorted the cards while reading them from the very beginning, others started sorting the cards after having read them, and still others read, for example, the first 20 cards quickly and sorted them roughly and then worked on the rest. Therefore, the time when they started reading the cards and the time they announced that they had completed the task were strictly timed with a stopwatch and recorded. The time taken for this was defined as the time participants took to complete the sorting task.

4.2.3 Data analysis

The collected data of L1 and L2 participants were analysed from two approaches. First, the time taken to complete the task (research question 1) and the number of clusters and words per cluster participant groups made (research questions 2 and 3) were analysed by means of unmatched *t*-tests. The analyses were done on the mean number of clusters and the mean number of words per cluster including single, isolated words, and then excluding those words. In the present sorting task participants were allowed to leave unsorted words as single, isolated words. It was expected that the presence of these words might affect the structures of the collected data. Thus, to examine the effect, it was necessary to analyse the results both with and without the single, isolated words. The mean number of single, isolated words and the mean largest cluster that participants made were also compared to each other between the two groups. This was done to investigate whether these variables had an impact on the sorting task results. Along with these *t*-test based analyses, L1 and L2 variability differences (research question 3) were further addressed by running *F*-tests on the results.

Second, a co-occurrence data matrix was constructed for both the L1 and L2 results, respectively. Word matrix analyses were made to further examine whether L1 lexical organisation has a larger number of clusters and a smaller cluster size (research question 2) while it is less varied than its L2 counterpart (research question 3). Taking into consideration the matrix analyses Meara and Schur (2002) carried out, the present matrix analyses were conducted regarding the mean number of (a) grouped words, (b) different words, (c) idiosyncratic associations (the ones only one participant made), and (d) associations that each word produces. These analyses were done by submitting the data to unmatched *t*-tests and *F*-tests as were in the analyses above. Regarding (a) the mean number of grouped words per cluster each word produced, the guideline Pollio (1963) developed was adopted. That is, for the results of each word in the matrix, a calculation is made on the number of other words it goes together with. Each of these different words is counted only once while ignoring how many participants made a link of that particular word with other words (see Meara & Schur, 2002, for the analysis).

4.3 Results

4.3.1 Time taken to complete task

To answer the first research question concerned with the time taken to complete the sorting task, Table 4.1 shows that the means and *SD*s of the time that the NS and NNS groups took to complete the task. On average, NNS took 3.56 minutes more to finish the task than NS did. An unmatched *t*-test revealed that there was a statistically significant difference between the two means (t (54) = 3.62, $p < 0.01$). Thus, the sorting task as a psycholinguistic data elicitation technique produced a similar result to Meara and Schur (2002) in that L2 participants took more time to complete the task than did their L1 counterparts.

Table 4.1. Time to complete 0.5K JACET sorting task

	NS (n = 28)	NNS (n = 28)	*t*-value	*F*-value
Mean	13.19	16.75	3.62**	0.61n.s.
SD	4.09	3.20		

Note. ** $p < 0.01$; n.s. = not significant.

4.3.2 Cluster number, size and variability

4.3.2.1 Between-group analysis

This section will address the between-group analyses of sorting tasks, examining the differences of L1 and L2 lexical organisations. The analyses are specifically done by means of unmatched *t*-tests and *F*-tests, while answering both research question 2 (concerned with cluster number and size) and research question 3 (regarding variability).

First, Table 4.2 shows the tabulation of the mean number of clusters participants made, where the count includes single, isolated words. The table shows that on average NS produced a slightly smaller mean cluster number than NNS did, but the difference did not reach a statistically significant level ((t (54) = 0.32, n.s.). As the *SD*s show, the variance of the NS results was larger than that of the NNS results. However, an *F*-test showed there was no statistically significant difference of variance between the two groups (F (27) = 0.73, n.s.).

Table 4.2. Mean number of clusters (which includes single, isolated words)

	NS (n = 28)	NNS (n = 28)	t-value	F-value
Mean	12.14	12.54	0.32n.s.	0.73n.s.
SD	4.87	4.16		

Note. n.s. = not significant.

To examine the effect of the single, isolated words on the results, the mean number of clusters participants made not counting single, isolated words was compared. The results are shown in Table 4.3. The mean cluster number that NS made was larger than NNS, but the difference did not reach a statistically significant level either ((t (54) = 0.66, n.s.). The variance of the NS results was smaller than that of the NNS results. However, an *F*-test revealed that there was no statistically significant difference of variance between the two groups (F (27) = 1.41, n.s.). Taking the results tabulated in Tables 4.2 and 4.3 into consideration, it was found that there was no statistically significant difference of mean cluster number and cluster variability between L1 and L2 participant groups. Thus, the present results turned out to be different from Meara and Schur (2002), which reported that L1 lexical organisation has a larger mean number of components and is less varied than its L2 counterpart.

Table 4.3. Mean number of clusters (which excludes single, isolated words)

	NS (n = 28)	NNS (n = 28)	t-value	F-value
Mean	9.11	8.54	0.66n.s.	1.41n.s.
SD	2.95	3.50		

Note. n.s. = not significant.

Second, Table 4.4 shows the tabulation of the mean word number per cluster participants made, where the count includes single, isolated words. The table shows that on average NS made a slightly larger mean cluster number than NNS did, but the difference did not reach a statistically significant level ((t (54) = 0.70, n.s.). The *SD* values show that there was larger individual variation between NS than NNS. This variation was not statistically significant (F (27) = 0.63n.s.), however.

Table 4.4. Mean number of words per cluster (which includes single, isolated words)

	NS (n = 28)	NNS (n = 28)	t-value	F-value
Mean	4.89 .	4.51	0.70n.s.	0.63n.s.
SD	2.24	1.78		

Note. n.s. = not significant.

To examine the effect of the single, isolated words on the results, the mean number of words per cluster not counting single, isolated words was compared. The results are displayed in Table 4.5. The table shows that the mean number of words per cluster the NS group made was smaller than the NNS group, but the difference did not reach a statistically significant level ((t (54) = 0.91, n.s.). The *SD* values show that there was smaller individual variance between NS than NNS. This difference was not statistically significant (F (27) = 1.87n.s.), however. Taking the results tabulated in Tables 4.4 and 4.5 into account, it was confirmed that there was no statistically significant difference in mean cluster size between L1 and L2 participant groups and variability. Thus, the present experiment produced different results from Meara and Schur (2002), which reported that L1 lexical organisation has a smaller mean number of components and is less varied than its L2 counterpart.

Table 4.5. Mean number of words per cluster (which excludes single, isolated words)

	NS (n = 28)	NNS (n = 28)	t-value	F-value
Mean	5.78	6.41	0.91n.s.	1.87n.s.
SD	2.16	2.95		

Note. n.s. = not significant.

In sum, all the results reported in this section showed that there was no substantial difference between L1 and L2 regarding the mean number of clusters, the mean cluster size and variability. While it was hoped that the results would provide evidence for Meara and Schur (2002), they failed to be statistically significant. Ultimately, the results were different from Meara and Schur (2002) who reported that L1 participants made more clusters, smaller cluster size and less variance than their L2 counterparts.

Related to these analyses, the mean number of single, isolated words (Table 4.6) and the mean largest cluster (Table 4.7) participants produced were tabulated.

Table 4.6. Mean number of single, isolated words

	NS ($n = 28$)	NNS ($n = 28$)	t-value	F-value
Mean	3.04	4.00	1.35n.s.	0.99n.s.
SD	2.69	2.67		

Note. n.s. = not significant.

Table 4.7. Mean largest cluster participants made

	NS ($n = 28$)	NNS ($n = 28$)	t-value	F-value
Mean	9.82	10.75	0.85n.s.	1.33n.s.
SD	3.79	4.38		

Note. n.s. = not significant.

According to Table 4.6, NS produced on average fewer single, isolated words than their L2 counterparts. Table 4.7 showed that NS produced on average the smaller "largest" cluster than their L2 counterparts. However, as the t-values show, these differences were not statistically significant, suggesting that these two parameters did not make a substantial impact on the results.

4.3.2.2 Co-occurrence matrix based analysis

In this section, analyses of the co-occurrence matrices (Appendices 4.2a and 4.2b) are made, and four relevant tables are provided to show the results. They are concerned with the mean number of (a) grouped words (Table 4.8), (b) different words (Table 4.9), and (c) idiosyncratic associations (Table 4.10). In addition, based on the matrices, (d) the number of participants who associated one word with another is tabulated in accordance with frequency (i.e., participant number) count based categorisation (Table 4.11). As stated earlier, this is done to further examine whether L1 lexical organisation has a larger number of clusters and a smaller cluster size while also being less varied than its L2 counterpart.

Table 4.8 shows the results of an examination into whether the number of grouped words per cluster that each word in the NS results makes is bigger than that of its NNS counterparts. It shows the means and *SD*s of clustered words each word made in the co-occurrence matrix (see Appendices 4.3a and 4.3b for the relevant data). The F-value reveals that the variances of the two groups were not equal ($F(49) = 2.655, p < 0.01$). Thus the Student's t-test for testing a statistically significant difference between the two means could not be adopted, and a Welch's t test was run instead. The results showed that the NS group produced a significantly smaller number of grouped words per cluster ($t(98) = 2.44, p < 0.05$). Thus, the mean number of grouped words per cluster each word made on the part of NS was a significantly smaller than that of NNS'. These results support Meara and Schur (2002) in that L1 participants produced a smaller component (cluster) number than L2 participants did, reflecting that L1 speakers are more aware of the semantic relations between words and how words may fall into distinct sets.

Table 4.8. Mean number of clustered words each word made in co-occurrence matrix ($k = 49$)

	NS ($n = 28$)	NNS ($n = 28$)	t-value	F-value
Mean	6.58	7.04	2.44*	2.66**
SD	0.70	1.14		

Note. * $p < 0.05$; ** $p < 0.01$.

Table 4.9 shows the mean number of different words each word produced. As stated above, following Pollio (1963) and Meara and Schur (2002), the calculation was done on the number of other words each tested word went together with in the co-occurrence matrix. For this specific analysis, how many participants made a link of the word with other words (strength of association) was ignored. The results are shown in Table 4.9.

Table 4.9. Mean number of other words in the cluster for each word ($k = 49$)

	NS ($n = 28$)	NNS ($n = 28$)	t-value	F-value
Mean	40.56	39.20	1.46n.s.	1.27n.s.
SD	4.37	4.93		

Note. n. s. = not significant.

Table 4.9 reveals that NS made a connection of slightly more words per word than NNS did, while there was less variance for NS than NNS. However, it was found that there was no statistically significant difference between the means (t (98) = 1.46, n.s.) nor the variance ($F(49) = 1.27$, n.s.). There was no substantial difference in the mean different words that a word produced between the two participant groups. This was distinctively different from Meara and Schur (2002), which reported the mean different words each word produced by L1 participants was significantly smaller than that of their L2 counterparts (8.5 different words for L1 and 12.4 for L2).

Table 4.10 was tabulated in order to address the effect of idiosyncrasy in the results. It contains the mean number of these behaviours that each word produced. The table reveals that L1 participants produced more idiosyncratic associations than L2 participants did. It also shows that the L1's *SD* was smaller than their L2 counterpart. However, the differences of the mean number and the variance were not statistically significant as the t-value (1.29n.s.) and the F-value (1.64n.s.) show. These results were also different from Meara and Schur (2002) which reported that L1 participants made statistically significantly less idiosyncratic associations (4.2) than did L2 participants (7.7).

Table 4.10. Mean number of idiosyncratic associations each word produced (k = 49)

	NS (n = 28)	NNS (n = 28)	t-value	F-value
Mean	11.20	10.32	1.29n.s.	1.64n.s.
SD	2.96	3.79		

Note. n. s. = not significant.

Table 4.11 contains the tabulation of the mean number of connections each word made with other words. The results were divided into four categories classifying different numbers of participants who made connections, where 28 (= number of participants in each group) was the potential maximum. That is, the results in the co-occurrence matrices were classified by means of how many participants made "zero" connections (i.e., the blank cells in the matrices where no participants made a word connection), and how many participants made a connection of a word with one to 10 other words, 11 to 20 and 21 to 28 (see Appendices 4.4a and 4.4b for the relevant data).

Table 4.11. Mean connection (participant) count tested words produced (k = 49)

No. of participants	NS (n = 28)		NNS (n = 28)		t-value	F-value
	Mean	*SD*	Mean	*SD*		
0	8.44	4.37	9.80	4.93	1.46n.s.	1.27n.s.
1-10	37.84	5.08	35.84	5.66	1.86n.s.	1.24n.s.
11-20	2.56	1.78	2.84	1.99	0.74n.s.	1.26n.s.
21-28	0.16	0.42	0.52	1.20	2.00*	8.08**

Note. n.s. = not significant, * $p < 0.05$; ** $p < 0.01$. The statistical analysis of the 21-28 participant number tabulation was done by a Welch's t test because of the unequal variances between the NS and NNS groups ($F(49) = 8.08, p < 0.01$.).

Table 4.11 reveals four salient features in the sorting task results. First, both groups predominantly produced the 1-10 participants category, with the NS mean being 37.84 and the NNS mean being 35.84. Second, concerned with the words producing "zero" cells in the matrix, NNS produced more cells of this sort than NS did. However, the difference failed to reach a statistically significant level (an unmatched t-test, t (98) = 1.46, n.s.). Third, in the classification of 11-20 participants, NNS produced a bigger mean value than NS did, but the difference failed to reach a statistically significant level (an unmatched t-test, t (98) = 0.74, n.s.). Last, while the value is extremely low, the NNS mean in the classification of 21-28 participants turned out to be statistically significantly larger than that of the NS's (an unmatched *t*-test, $t(98) = 2.00$, $p < 0.05$.). It should be noted that the significant difference between the L1 and L2 results did not have a substantial impact on the overall results, considering the extremely low means for both groups (0.16 for NS and 0.52 for NNS).

In sum, by this co-occurrence matrix based analyses, it was found that with L1 participants, only a small portion of the number of connections each word made with other words was substantially smaller than that of their L2 counterparts. Regarding the mean number of different words each word produced and the idiosyncratic responses, there were no meaningful differences between the two participant groups.

4.4 Discussion

The results reported in section 4.3 showed that the sorting task in the present research design did not necessarily produce distinctive L1 and L2 differences in underlying lexical organisations. This is especially true with between-group analyses. L1 participants tended to produce a larger number of clusters and a smaller cluster size than L2 participants did. This was the same results as Meara and Schur's (2002) study. However, regarding all the analyses including the mean largest cluster produced, there were no statistically significant differences between the two groups. Meanwhile, when co-occurrence matrix based analyses were done, it was found that L1 participants produced a significantly smaller number of grouped words per cluster. It was also revealed that the number of associated words that 21-28 NS participants produced was significantly smaller than its L2 counterpart. Taking all these results into account, we can argue that some word specific factors might have accounted for the weaker than expected ability to tap into L1 and L2 differences in sorting behaviour. This is the most crucial point to be examined regarding the validity of the present first sorting task in this research project. It also seems to be worthwhile to examine if there are any qualitative differences for the sorting results between L1 and L2 participants, and if there are, what exactly they are. Furthermore, the results of the present experiment were similar to Meara and Schur's (2002) study in that L2 participants took significantly more time to complete the task.

In the next section, I will begin by discussing this task completion time issue in the sorting task, and then go on to the issue of the task sensitivity and tested words. Lastly, I will explore the qualitative differences between the groups concerned with the salient association pairs (the ones where a substantial number of participants produced in the sorting task) in the clusters participants made.

4.4.1 Task completion time

It was revealed that in the sorting task L1 participants took an average of 13.19 minutes and their L2 counterparts took an average of 16.75 minutes. The 3.56 minute difference was statistically significant. Although it is too early to draw a conclusion, the present results suggest that the sorting task is likely to be more demanding and time-consuming to L2 participants than to L1 participants. In Meara and Schur (2002) in which they implemented a restricted word association test, L2 participants took more time to complete the task than L1 participants did. Meanwhile, in Wilks and Meara (2002), which used another WAT experiment, L2 participants took less time than their L1 counterparts did. Examining these results together, the present sorting task is more similar to Meara and Schur's (2002) restricted WAT than Wilks and Meara's (2002) restricted WAT in task demand and completion time. This can be explained by the task directions the two studies gave to their participants. In Meara and Schur, participants were directed to select another verb from a set of 50 verbs that was most associated with it. In Wilks and Meara, participants were instructed to read a set of five words (40 sets in total) and circle any two words in it that they thought were associated. Thus, the number of words (in these cases five or 50 words) participants need to work out the semantic relationships of seems to play a major role in relation to how long it took participants to complete the tasks.

Moreover, this difference in the tasks affected L2 participants more significantly than it did L1 participants. In other words, regardless of whether we are dealing with a restricted WAT or a sorting task, when participants are required to work out the semantic relationships among a large number of words, they need to process the same number of possible links between them. Faced with a large number of 50 words to process to complete the task, L2 participants were put into more challenging situations. In processing a large number of L2 words and the relationships between them, they were likely forced to activate their L2 lexical knowledge as well as their conceptual knowledge and L1 lexical knowledge. This was not the case with L1 participants, who did not need to rely on L2 lexical knowledge during task completion. Ultimately L2 participants took more time to complete the present sorting task and the restricted WAT Meara and Schur (2002) conducted than their L1 counterparts. This was not the case with the less demanding Wilks and Meara's (2002) experiment where L2 participants took an average of 10.38 minutes and L1 participants took an average of 11.25 minutes. This would "seem to confirm that they [L2 participants] were not struggling with unknown words and that the task of identifying associations between words was a relatively spontaneous one" (Wilks & Meara, 2002, p. 318). To be more precise, however, the sorting task lies somewhat between Wilks and Meara (2002) and Meara and Schur (2002) in task demand and completion time. This is because in Meara and Schur (2002), L1 participants took 10 to 15 minutes and L2 participants needed "a full class period to complete the task" (p. 170). The present sorting task took far less time for L2 participants to complete it. Thus a distinctive merit of this sorting task is that participants tapped into their L2 lexical organisation while needing less time and cognitive effort.

While it seems highly plausible that in the sorting task L1 participants took less time than L2 participants, we should refrain from drawing that conclusion at this stage of the research project. This is because the tested words in Meara and Schur (2002), Wilks and Meara (2002) and the present experiment are different from each other and therefore difficult to compare to each other reliably. In Meara and Schur, they used 50 verbs selected from Nation's (1986) first frequency band. In Wilks and Meara, they used 40 sets of 5 words "randomly chosen from the Français Fondamental list: approximately the first 1000 most frequency words in French excluding grammatical items (Gougenheim et al., 1956)" (p. 314). Thus, both of the restricted WAT studies selected tested words by means of different sources and criteria from each other as well as the present sorting task experiment. The most distinct difference of the present experiment from the other two studies is that for the present experiment I randomly selected 50 words from among the most frequent 500 words in the JACET 8000. Accordingly, there were grammatical items (e.g., one pronoun, one determiner, and one conjunction) among the tested words. Furthermore, the ratio of non-grammatical items was heavily skewed: 30 nouns (60%), five verbs (10%), six adjectives (10%) and six adverbs (12%). To reach a conclusion on the L1 and L2 difference in task completion time, I will need to implement a series of sorting tasks where the tested words are selected by means of different, but consistent criteria. (The issue of grammatical items among the tested words will be discussed later in this chapter.)

4.4.2 Task sensitivity and the tested words

In this section, I will address the issue of why the present sorting task did not produce results showing distinctive L1 and L2 differences in lexical organisation. This is particularly concerned with examining the reason why L1 participants failed to make a larger mean number of clusters and a smaller mean cluster size, both of which had not been predicted. As confirmed earlier, the mean number of word clusters NS made was slightly larger than NNS and the size of grouped words per cluster NS made was somewhat smaller than NNS. In this regard, the sorting task seems to have potential for tapping into the L1 and L2 organisational differences confirmed by Meara and Schur (2002). The problem is that the results did not reach a statistically significant level in terms of between-group analyses. Considering these results as a whole, I hypothesised that the low sensitivity of the present sorting task can be attributed to the features of the tested words in the experiment. Accordingly, this issue will be discussed in view of the impact of the tested words on the overall sorting results. Two main points will be addressed.

First, it should be noted that only a small portion of the tested words had the sensitivity to produce L1 and L2 differences in the sorting task and the effect on the overall results was limited. This can be seen in Table 4.11, which contains the calculation of the mean number of connections each word made with other words (see 4.3.2.2 Co-occurrence matrix based analysis). The table shows that of the four classifications tested, a statistically significant difference in the number of connections between the two groups was found only in the classification of 21-28 connections (participants). Furthermore, the number of tested words that produced a meaningful between-group difference was extremely low, being less than one word on average for both groups (0.16 for L1 and 0.52 for L2). Meanwhile, the dominant number of connections a word made fell into the 1-10 connections category for both groups (L1 = 37.84 and L2 = 35.84 words on average), followed by the "zero" connections category (L1 = 8.44 and L2 = 9.80 words), and the 11-20 connections category (L1 = 2.56 and L2 = 2.84 words). All of these results revealed no statistically significant differences between the NS and NNS groups. Thus, it was found that an extremely small number of tested words had the sensitivity to distinguish L1 participants from their L2 counterparts in the sorting task.

Second, related to the task sensitivity issue discussed above, I will examine which words produced statistically significant L1 and L2 differences in the sorting task results. Based on the information of the co-occurrence matrices (Appendices 4.2a and 4.2b), Table 4.12 summarises average numbers of clusters each word produced (counting includes the word in question) and the difference between the NS and NNS groups in descending order of *t*-values. Regarding the *t*-values, Table 4.12 shows what happens if we adopt a 10% significance level, rather than the more conventional 5% level. This is done because, as confirmed above, only a small portion of the tested words produced statistically significant differences between the two groups. To draw a wider picture of how well all the 50 tested words contributed to the overall results, I decided that it would be worthwhile to detect the words that almost brought about statistically significant differences as well.

Table 4.12. Average number of clusters each word produced (k = 50)

word	Average		NS-NNS	*t*-value	*F*-value
	NS	NNS			
president	6.36	8.75	-2.39	2.15*	**2.78****
also	5.71	3.61	2.11	1.73*	0.48
police	6.39	8.29	-1.89	1.69*	**2.58***
war	7.11	9.00	-1.89	$1.64^{\dagger}$	**2.36***
dream	6.14	7.79	-1.64	$1.59^{\dagger}$	1.51
law	6.71	8.54	-1.82	$1.58^{\dagger}$	**2.84***
history	7.11	8.57	-1.46	$1.43^{\dagger}$	**3.16***

(*table continues on next page*)

Table 4.12. (*continued*)

arm	5.54	7.46	-1.93	1.42†	**2.62***
arrive	6.29	7.82	-1.54	1.38†	1.58
walk	6.07	7.50	-1.43	1.38†	1.24
very	5.21	3.64	1.57	1.35†	**0.40***
power	7.18	8.64	-1.46	1.32†	**2.72***
boy	5.11	6.21	-1.11	1.28	1.95
close	6.25	7.57	-1.32	1.20	1.15
open	6.07	7.36	-1.29	1.19	1.74
find	6.11	7.18	-1.07	0.94	1.24
figure	6.29	7.50	-1.21	0.93	1.62
place	7.46	6.50	0.96	0.92	0.65
lot	6.14	5.11	1.04	0.91	0.60
doctor	5.86	6.57	-0.71	0.75	1.67
step	5.75	6.46	-0.71	0.74	1.45
century	6.93	7.68	-0.75	0.71	**3.89****
business	6.82	7.50	-0.68	0.70	**2.99****
cry	6.39	7.14	-0.75	0.67	1.00
stand	6.64	7.39	-0.75	0.62	1.05
keep	6.32	7.11	-0.79	0.59	0.93
believe	7.61	7.04	0.57	0.57	1.71
ago	7.71	8.32	-0.61	0.55	1.29
shop	6.54	7.11	-0.57	0.55	0.70
nature	6.29	5.64	0.64	0.55	1.96
help	6.89	7.57	-0.68	0.49	1.39
understand	7.71	7.14	0.57	0.45	**2.33***
while	6.57	6.07	0.50	0.44	0.49
country	7.64	8.04	-0.39	0.35	**2.42***
area	7.61	7.25	0.36	0.34	0.75
then	7.43	7.79	-0.36	0.32	1.05
nothing	5.54	5.14	0.39	0.32	2.11
matter	6.43	6.86	-0.43	0.31	1.18
clear	6.54	6.21	0.32	0.3	1.11
form	6.75	7.11	-0.36	0.26	1.22
person	6.36	6.07	0.29	0.25	0.70
next	7.32	7.57	-0.25	0.24	0.55
already	7.39	7.14	0.25	0.21	0.98
all	6.18	5.93	0.25	0.19	1.31
street	7.07	7.25	-0.18	0.17	0.68
air	5.46	5.64	-0.18	0.17	**2.97***
social	7.89	8.11	-0.21	0.16	1.67
our	5.93	6.14	-0.21	0.14	1.34
reason	7.32	7.21	0.11	0.09	1.82
early	6.93	6.96	-0.04	0.04	**0.38***

Note. Results are shown in descending order of *t*-values. $^{\dagger} < 0.10$; * $p < 0.05$; ** $p < 0.01$. Calculations whose *F*-values are highlighted in bold were done by Welch's *t* tests because of unequal variances between the NS and NNS groups.

Table 4.12 reveals that there was a statistically significant difference (unmatched t-tests) found between NS and NNS regarding the production of a small number of clusters when the alpha level is set at 10% for 10 words: *president, police, war, dream, law, history, arm, arrive, walk* and *power*. When the alpha level is strictly set at 5% or lower, however, the number of statistically significant words drops sharply to just two (*president* and *police*). Thus, only these two words out of the 50 words had the sensitivity to substantially distinguish L1 participants from L2 participants, reflecting more awareness by the native speakers of the semantic relations between words and how they may fall into distinct sets. On the contrary, there were two words (*also* and *very*) whose L2 results were significantly lower than the L1's. These words were not good choices for the sorting task, considering the counterintuitive results. Another problem lies in the fact that 14 (28.0%) tested words shown in bold in Table 12 produced statistically significant F-values between the results of L1 and L2 participants. This suggests that with these words, individual variances in the results were very large and the subsequent between-group analysis was problematic.

Taking the analyses above into account, the low sensitivity of the present sorting task is attributed to the way the tested words were selected. The 50 words were selected randomly from among the most frequent 500 words in the JACET 8000. There were two problems in choosing the words for the task in this way. One is concerned with the issue of using extremely high frequency words in the experiment. That is, because the words were selected from the very high frequency of 500 words, they seemed to have little power to distinguish L1 speakers from L2 speakers in lexical organisation. These high frequency words had been learned at the junior/senior high school levels in Japan and had been integrated into the L2 mental lexicons of the advanced level adult L2 participants. This suggests that in future experiments, tested words should not be such extremely high frequency words as the 50 words used in the present experiment.

Another problem derived from selecting the tested words randomly for the experiment. Some of the words appeared to have puzzled participants and led them to carry out the task in inconsistent ways. This was observed in the counterintuitive results of *also* and *very* whose L2 results were significantly lower than the L1's when compared to the overall tendency. These grammatical items were not appropriate and should not have been included in the tested words.

In sum, it was found that the present sorting task had only a small number of tested words that had the sensitivity to tap into L1 and L2 differences by means of the sorting task. The main cause of the problem was believed to lie in the random selection of very high frequency words for the task. These are the points to be reconsidered in future experiments to get more reliable data in investigating L1 and L2 differences in cluster structures of the mental lexicons.

4.4.3 Qualitative differences of word combinations in clusters

In this section, I will explore the qualitative differences between L1 and L2 participants regarding the cluster structures that the sorting task tapped into. For the analysis, salient association pairs (the ones which a substantial number of participants produced in the sorting task) in clusters were searched for in the L1 and L2 co-occurrence matrices (Appendices 4.2a and 4.2b). The cut-off point was set at 15, meaning that I identified association pairs in the matrices that more than half of the participants (i.e., 15 or more participants out of the total 28) produced (e.g., 19 L1 participants produced the *ago-already* link while 21 L2 participants produced it). The results of the analysis are shown in Appendix 4.5a for the NS group and Appendix 4.5b for the NNS group. This post hoc analysis needs to be treated with caution since there is considerable overlap between identified linked words for each group. However, the analysis reveals the overall tendency of the sorting results to support the validity of the present sorting task. Table 4.13 shows the summary of associations of linked words that 15 or more participants produced.

Table 4.13. Associations of linked words having a substantial number of productions (ones that 15 or more participants produced)

NAME OF	NS ($n = 28$)		NS ($n = 28$)	
ASSOCIATION	Frequency	%	Frequency	%
TIME	21	30.9	29	33.7
PLACE	12	17.6	8	9.3
THOUGHT	10	14.7	6	7.0
POWER	7	10.3	14	16.3
HISTORY	6	8.8	3	3.5
MOVEMENT	4	5.9	6	7.0
PERSON	2	2.9	8	9.3
OPPOSITE WORDS	2	2.9	6	7.0
NATURE	2	2.9	2	2.3
OTHERS	2	2.9	4	4.7
TOTAL	68	100%	86	100%

Table 4.13 shows that the randomly selected 50 words in the present experiment were not randomly sorted, but were sorted in a consistent way by means of the semantic relatedness both L1 and L2 participants identified. The TIME category (21 (30.9%) association pairs for NS and 29 (33.7%) for NNS) accounts for the largest portion that both groups produced.

PLACE (12 (17.6%) for NS and 8 (9.3%) for NNS), THOUGHT (10 (14.7%) for NS and 6 (7.0%) for NNS) and POWER (7 (10.3%) for NS and 14 (16.3%) for NNS) associations were also produced a considerable amount of times by both groups. There is another piece of evidence for the general tendency of participants to sort the words consistently by means of semantic relatedness. The participants sorted the words belonging to different parts of speech into the same clusters as long as they went together according to meaning. For example, L1 participants produced such pairs as *ago* (adverb) - *century* (noun), *already* (adverb) – *while* (conjunction), and *also* (adverb) - *while* (conjunction). Meanwhile, L2 participants produced such pairs as *already* (adverb) – *while* (conjunction), *early* (adverb) – *while* (conjunction), and *social* (adjective) – *law* (noun). (See Appendices 4.5a and 4.5b for details.) These results also indicate that the cluster structures of L1 and L2 lexical organisations are semantically different from each other and the sorting task taps into these underlying structures. Considering the potentials of the sorting task to probe into the qualitative differences of L1 and L2 lexical organisations, I argue that it is worth carrying out further experiments using the task under different research designs to address aspects of the cluster structures.

Taking a closer look at the linked words that 15 or more participants produced in Appendices 4.5a and 4.5b, it was found that there are association pairs that were produced only by a significant number of participants in the NS group. They were *doctor-police*, *shop-business*, *understand-reason*, *step-walk* and *also-while*. Meanwhile, the following are the combinations of words that 15 or more L2 participants produced but their L1 counterparts did not: *war-law*, *social-law*, *find-believe*, *find-understand*, *matter-reason* and *cry-help*. Furthermore, it should be noted that NS tended to produce *understand-reason* while NNS produced *matter-reason* in large number. These results are also evidence for the fact that there are in fact qualitative differences of linked words between L1 and L2 lexical organisations.

4.5 Conclusion

In this chapter, I reported the first experiment I did using a sorting task that was developed, while considering the limitations of the sorting tasks Haastrup and Henriksen (2000) implemented. I selected the 50 words from the most frequent 500 words in the JACET 8000 so that the participants knew them all and the task tapped into their lexical organisation rather than their lexical knowledge. This word selection was necessary to examine the underlying cluster structures of the L1 and L2 mental lexicons that should have been reflected in the sorting task results. A problem that manifested, however, was that the very high frequency words selected had less sensitivity to tap into the lexical organisations than expected. These very high frequency words were integrated fully into the L2 mental lexicons of the participants and thus one could not clearly distinguish L1 participants from L2 participants in the sorting task results.

The results tended to show that L1 lexical organisation has a larger number of clusters, a smaller cluster size and is less varied than L2 lexical organisation. But many confirmed differences failed to reach statistically significant levels. Most notably, the L1 and L2 differences in the mean clustered words each word made and the mean number of connections each word made with other words regarding only the tested words 21-28 participants produced turned out to be statistically substantial. These results suggest that future sorting tasks should adopt a different word selection method from the present one. The method of selecting words randomly for the experiment from a word list should also be reconsidered.

The present experiment showed that a sorting task as a psycholinguistic data elicitation task is similar to the restricted word association test Meara and Schur (2002) implemented in terms of the task demand on participants' cognitive effort and time consumption. This particularly seems to affect L2 participants more than their L1 counterparts. It was also revealed that both L1 and L2 participants worked on the sorting task consistently by means of semantic relatedness, whereas the tested words were selected randomly. While the evidence is limited, the present analyses found that L1 lexical organisation is qualitatively different from its L2 counterpart in terms of cluster structures.

In Chapter 5, I will revise the present sorting task, considering the limitations identified above, and further address the issues of L1 and L2 differences in cluster numbers, size and variability.

Chapter 5: Native-like links in cluster structures

5.1 Introduction

In Chapter 4, I developed the first sorting task of my own and with the task implemented an experiment to examine the cluster structures of L1 and L2 mental lexicons. Unfortunately, the overall results were more discouraging than encouraging. That is, the task did not produce prominent L1 and L2 differences. L1 participants tended to produce a larger number of clusters, a smaller cluster size and they were less varied than their L2 counterparts, but many of the differences did not reach statistically significant levels.

However, there were also promising results. It was shown that participants in both groups consistently sorted the words into clusters by ways of meaning. Thus, it was confirmed that the sorting task played a reliable role to tap into the cluster structures of lexical organisation. When co-occurrence matrix analyses were run, there were five native-like associations (i.e., links), namely *doctor-police*, *shop-business*, *understand- reason*, *step-walk* and *also-while* that more than half of L1 participants produced but that less than half of their L2 counterparts did. Matrix analyses also revealed that the mean number of grouped words per cluster each word made on the part of NS was significantly smaller than that of NNS. Moreover, L1 and L2 differences in the mean number of connections each word made with other words regarding the tested words that 21-28 participants produced proved to be statistically significant.

As Meara (2004) indicated after comprehensively reviewing research on the structures of the mental lexicon, most applied linguists agree that "vocabularies are not just collections of words, and that vocabularies are essentially interlocking networks" (p. 137). The results of the previous experiment confirmed this claim that lexical organisation is essentially structured firmly together. However, it should be noted that these results were too limited to explain L1 and L2 differences of cluster structures of overall lexical organisation. Given this, it becomes worthwhile to revise the sorting task so that it taps into the differences more clearly and an experiment can be done with it.

The problem in the sorting task reported in the previous chapter lay in the word selection procedure. The 50 tested words were randomly selected from among the most frequent 500 words in the JACET 8000. Closer analysis of those words revealed that only two words (*president* and *police*) out of the 50 words had the sensitivity to substantially distinguish L1 participants from their L2 counterparts. Most of the very high frequency words appeared to have already been integrated into the L2 lexical organisation of NNS participants and produced statistically non-significant results between the NS and NNS groups. This suggests that the revised sorting task should use less high frequency words that NNS still know the meanings of. In addition, the method of selecting tested words randomly from a word list was changed for the revised sorting task. The sorting task reported in the previous chapter contained words (i.e., adverb *also* and adjective *very*) whose results were counterintuitive and contradicted the results that other words produced. Moreover, the proportion of word classes for tested words was highly skewed: 30 (60%) nouns, five (10%) verbs, six (12%) adjectives, six (12%) adverbs, one (2%) pronoun, one (2%) determiner and one (2%) conjunction. This skewed proportion of the type of tested words which was generated by random selection seems to have affected participants' sorting behaviour and produced rather discouraging results. Thus, for the revised sorting task, tested words were selected from a single word class to obtain more discernable L1 and L2 differences.

By making the revisions above, native-like links in L1 and L2 cluster structures were examined. In addition, the research questions posed in the previous chapter will be further addressed. More specifically, the task completion time issue was addressed to confirm whether L1 participants actually take less time than L2 participants as the previous sorting task showed (research question 1a). Which sorting task (the previous or revised) requires participants more time to finish is also discussed (research question 1b). Moreover, the issue of cluster number, size and variability is addressed (research question 2a). This issue is also investigated by looking at the differences in the results that the previous sorting task and the present one produce (research question 2b). Furthermore, I will attempt to answer whether L1 participants will produce a significantly larger number of distinctively different associations from their L2 counterparts in terms of native-like links in the task results (research question 3). These points are summarised below.

1a. Do L1 participants take less time to complete a sorting task than L2 participants?
1b. Which sorting task takes participants more time to complete, the previous one or the revised one?
2a. Do L1 participants make a larger number of clusters and fewer words per cluster than their L2 counterparts do? Is L1 lexical organisation less varied than its L2 counterpart?
2b. Regarding the variables in 2a, are there distinctive differences between the results produced by the previous sorting task and the present one?
3. Do L1 participants produce a larger number of native-like links than their L2 counterparts?

5.2 Method

5.2.1 Participants

In this study, there were two participant groups. The first was comprised of 30 adult, native speakers of English (NS). They were teachers of English in Kumamoto, visiting scholars and students studying at Kumamoto University. The second consisted of 30 adult, advanced-level Japanese speakers of English (NNS). They were either English teachers at the college level or persons having a high competence of English as judged by a TOEFL score of 213 or more on the computer-based version or a score of 550 or more on the paper-based version or a TOEIC score of 730 or more that had been taken within the last two years.

5.2.2 Data collection

I randomly selected 50 verbs from verbs that are among the first, 1000 high frequency words from the JACET 8000: *to accept, to add, to agree, to argue, to ask, to avoid, to begin, to believe, to build, to buy, to carry, to choose, to consider, to create, to decide, to describe, to discover, to discuss, to enter, to expect, to explain, to find, to follow, to get, to give, to grow, to hear, to imagine, to improve, to introduce, to join, to learn, to listen, to lose, to meet, to prepare, to protect, to read, to receive, to refuse, to remember, to sell, to speak, to spend, to suggest, to teach, to tell, to throw, to wait,* and *to write.*

I decided to use verbs in particular for the present experiment because verbs are usually the most difficult ones for non-native speakers of English to master. Examining the errors committed by advanced German learners of English in speech, Lennon (1996) reported that verb errors accounted for 13% of all errors, and in most cases they were related to simple high frequency verbs. He indicates that "learners may have a broad outline of verb meaning, but that their lexical knowledge is hazy concerning polysemy, contextual and collocational restrictions, phrasal verb combinations, [and] grammatical environment" (p. 35). This would likely be the case with the NNS participants (i.e., the advanced Japanese learners of English in this experiment) as well. It is predicted that L1 and L2 differences would be the most evident in the results of the sorting task using high frequency verbs, reflecting the fact that NSs are more aware of differences between verbs than NNSs are.

After piloting, the task was administered to each of the participants individually. Each pile of 50 cards, on which the individual English words were printed, was shuffled and bound with a rubber band and put into an envelope for each participant. The order of cards was fixed so that participants would read them in the same way. The directions for the task were the same as those for the sorting task reported in Chapter 4, except that since the tested words were verbs, participants were invited to sort them into groups of words that they thought would go together according to meaning. The directions were written in English for native speakers of English, and in Japanese for Japanese participants.

5.2.3 Data analysis

The present experiment was closely related to the previous one reported in Chapter 4. Therefore, major analyses were similarly made regarding the mean task completion time, mean number of clusters and mean number of words per cluster the NS and NNS groups produced. The mean numbers of single, isolated words and the mean largest cluster that participants made were also compared to each other. Unmatched Student's *t*-tests were run to test whether the means were statistically distinctive from each other. I also ran unmatched *t*-tests on the data produced by the previous sorting task and the present one, searching for any differences they might have generated. This was done to examine whether revising the task had an effect on clearly tapping into L1 and L2 differences in lexical organisation. These analyses are all related to research questions 1a, 1b, 2a and 2b. To answer research question 3 addressing native-like links in cluster structures, I elicited pairs of "native-like" links in the clusters participants made. The elicitation was made on an L1 co-occurrence matrix by identifying pairs of links that half (= 15) or more of NS participants produced. Using the identified pairs as the baseline of native-like links, I located them in individual participant's results of each group. Then I ran an unmatched *t*-test on the results to examine whether there was a significant difference between the two groups regarding native-like links.

5.3 Results

5.3.1 Time taken to complete task

This section will answer research questions 1a and 1b, which both address L1 and L2 differences in task completion time. Table 5.1 shows the means and *SD*s of the amount of time that the NS and NNS groups took to complete the present sorting task using 50 high frequency verbs. On average, NNS took 3.04 minutes more to finish the task than NS did. An unmatched *t*-test revealed that there was a statistically significant difference between the two means ($t(58) = 3.24$, $p < 0.01$). Thus, the revised sorting task produced a similar result to the previous sorting task in that L2 participants took more time to complete the task than their L1 counterparts did. This confirmed that a sorting task is more similar to Meara and Schur's (2002) limited word association test than to Wilks and Meara's (2002) in that L2 participants took more time to complete the task than their L1 counterparts did.

Table 5.1. Time to complete 1K JACET verbs sorting task

	NS ($n = 30$)	NNS ($n = 30$)	*t*-value	*F*-value
Mean	14.76	17.80	3.24**	0.72n.s.
SD	3.91	3.32		

Note. ** $p < 0.01$; n.s. = not significant.

Addressing whether the revisions of the sorting task had an effect on task completion time, Table 5.2 shows that L1 participants took 1.57 minutes more to complete the revised sorting task using randomly selected 1K (i.e., the most frequent 1000) verbs than the previous one using 0.5K (i.e., the most frequent 500) words. The number of participants in the present experiment (n = 30) was different from the previous experiment (n = 28) and so the data could not be directly compared. Therefore, unmatched Welch t tests were run on the data to examine whether the means were distinctively different.

Table 5.2. Sorting task completion time for NS: 0.5K JACET words vs. 1K JACET verbs

	0.5K (n = 28)	1K (n = 30)	t-value	F-value
Mean	13.19	14.76	1.50n.s.	1.10n.s.
SD	4.09	3.91		

Note. 0.5K = the sorting task using 50 words randomly selected from the most frequent 500 words in the JACET 8000; 1K = the sorting task using 50 words randomly selected from the most frequent 1000 verbs in the JACET 8000. The same notations will be used hereafter in this chapter. n.s. = not significant.

Meanwhile, Table 5.3 shows that L2 participants took 1.05 minutes more to complete the revised sorting task. The results revealed that there was no statistically significant difference in task completion time regarding both the NS group ($t(58) = 1.50$, n.s.) and the NNS group ($t(58) = 1.23$, n.s.). Thus, although high frequency verbs tended to require participants more time to sort than randomly selected very high frequency words did for the NS and NNS groups, the differences were not statistically different.

Table 5.3. Sorting task completion time for NNS: 0.5K JACET words vs. 1K JACET verbs

	0.5K (n = 28)	1K (n = 30)	t-value	F-value
Mean	16.75	17.80	1.23n.s.	0.93n.s.
SD	3.20	3.32		

Note. n.s. = not significant.

5.3.2 Cluster number, size and variability

This section will answer research questions 2a and 2b, which are concerned with cluster number, size and variability. As in the previous section, analyses will first be made on L1 and L2 comparisons regarding the present sorting task, followed by between-task comparisons of the previous sorting task and the present one. For this purpose, four sets of tables (mean number of clusters, mean number of words per cluster, mean number of single, isolated words, and mean largest cluster participants made) will be shown. Each of the tables has three related tables.

Table 5.4 shows the mean number of clusters participants made in the present sorting task, where the count excludes single, isolated words. The table reveals that on average NS generated a slightly larger mean cluster number than NNS did, but the difference was not a statistically significant one ($t(58) = 0.72$, n.s.). As the *SD*s show, the variance of the NS results was smaller than that of the NNS results. However, an *F*-test revealed that there was no distinctive difference in variability between the two groups ($F(29) = 1.25$, n.s.). Thus, the present sorting task failed to distinctively distinguish NS from NNS in mean cluster number and the variability that participants generated.

Table 5.4. Mean number of clusters (which excludes single, isolated words)

	NS ($n = 30$)	NNS ($n = 30$)	*t*-value	*F*-value
Mean	9.93	9.33	0.72n.s.	1.25n.s.
SD	3.05	3.41		

Note. n.s. = not significant.

Table 5.5 shows that L1 participants produced on average a slightly larger number of clusters with the sorting task using 1K verbs than they did with the task using 0.5K randomly selected words. Similarly, Table 5.6 shows that L2 participants also produced on average a slightly larger number of clusters with the sorting task using 1K verbs than they did with the task using 0.5K randomly selected words. Unmatched Welch *t* tests revealed that there was no statistically significant difference in the mean number of clusters for both the NS group ($t(58) = 1.05$, n.s.) and NNS group ($t(58) = 0.88$, n.s.). The sorting task using 1K verbs was likely to produce a larger number of clusters than the task using 0.5K random words for the two groups, but both differences did not reach a statistically significant level.

Table 5.5. Mean number of clusters NS produced (which excludes single, isolated words): 0.5K JACET words vs. 1K JACET verbs

	0.5K ($n = 28$)	1K ($n = 30$)	*t*-value	*F*-value
Mean	9.11	9.93	1.05n.s.	0.93n.s.
SD	2.95	3.05		

Note. n.s. = not significant.

Table 5.6. Mean number of clusters NNS produced (which excludes single, isolated words): 0.5K JACET words vs. 1K JACET verbs

	0.5K ($n = 28$)	1K ($n = 30$)	t-value	F-value
Mean	8.54	9.33	0.88n.s.	1.06n.s.
SD	3.50	3.41		

Note. n.s. = not significant.

Table 5.7 shows the mean number of words per cluster participants in the present sorting task made, where the count excludes single, isolated words. The table reveals that on average NS generated a slightly smaller mean number of words per cluster than NNS did by, but the difference did not reach a statistically significant level ($t(58) = 0.28$, n.s.). As the *SD*s show, the NS's variance was somewhat larger than that of their NNS counterparts. However, an F-test revealed that there was no substantial difference in variability between the two groups ($F(29) = 0.65$, n.s.). Thus, the present sorting task failed to clearly distinguish NS from NNS in terms of the mean number of words per cluster and the variability that participants produced.

Table 5.7. Mean number of words per cluster (which excludes single, isolated words)

	NS ($n = 30$)	NNS ($n = 30$)	t-value	F-value
Mean	5.39	5.57	0.28n.s.	0.65n.s.
SD	2.75	2.21		

Note. n.s. = not significant.

Table 5.8 shows that L1 participants produced on average a slightly smaller number of words per cluster with the sorting task using 1K verbs than they did with the task using 0.5K randomly selected words. Meanwhile, Table 5.9 shows that L2 participants likewise produced on average a smaller number of words per cluster with the sorting task using 1K verbs than they did with the task using 0.5K randomly selected words. Unmatched Welch t tests revealed that there was no statistically significant difference in the mean number of clusters for either the NS group ($t(58) = 0.61$, n.s.) or the NNS group ($t(58) = 1.24$, n.s.). The sorting task using 1K verbs had a tendency to produce a smaller number of words per cluster than that of the sorting task using 0.5K random words for the two groups, but both differences were not statistically significant ones.

Table 5.8. Mean number of words per cluster NS produced (which excludes single, isolated words): 0.5K JACET words vs. 1K JACET verbs

	0.5K ($n = 28$)	1K ($n = 30$)	t-value	F-value
Mean	5.78	5.39	0.61n.s.	0.62n.s.
SD	2.16	2.75		

Note. n.s. = not significant.

Table 5.9. Mean number of words per cluster NNS produced (which excludes single, isolated words): 0.5K JACET words vs. 1K JACET verbs

	0.5K (n = 28)	1K (n = 30)	t-value	F-value
Mean	6.41	5.57	1.24n.s.	1.78n.s.
SD	2.95	2.21		

Note. n.s. = not significant.

Analyses of between-group and between-task comparisons regarding the mean number of single, isolated words, and the mean largest cluster participants produced were tabulated and are shown in Table 5.10 through Table 5.15 below. These analyses were made in order to investigate the effect of the types of words used on the overall results of each of the sorting task.

Table 5.10 shows the mean number of single, isolated words participants produced in the revised sorting task. The table reveals that on average NS generated a smaller mean number of single, isolated words than NNS did, but the difference was not a statistically significant one ($t(58) = 0.85$, n.s.). Thus, the single, isolated words did not have a significant impact on the overall results and failed to generate any differences between the NS and NNS groups.

Table 5.10. Mean number of single, isolated words

	NS (n = 30)	NNS (n = 30)	t-value	F-value
Mean	3.63	4.60	0.85n.s.	1.76n.s.
SD	3.76	4.99		

Note. n.s. = not significant.

Table 5.11 shows that NS produced on average a slightly larger mean number of single, isolated words with the sorting task using 1K verbs than with the task using 0.5K randomly selected words. A similar result was gained in the case of NNS where participants produced on average a larger number of single, isolated words with the sorting task using 1K verbs (Table 5.12). However, as the t-values show, these differences did not reach a statistically significant level. This suggests that the revised sorting task was not different from the previous one in that both tasks produced broadly similar results in regard to the mean number of single, isolated words the two participant groups generated.

Table 5.11. Mean number of single, isolated words NS produced: 0.5K JACET words vs. 1K JACET verbs

	0.5K ($n = 28$)	1K ($n = 30$)	t-value	F-value
Mean	3.04	3.63	0.69n.s.	0.51n.s.
SD	2.69	3.76		

Note. n.s. = not significant.

Table 5.12. Mean number of single, isolated words NNS produced: 0.5K JACET words vs. 1K JACET verbs

	0.5K ($n = 28$)	1K ($n = 30$)	t-value	F-value
Mean	4.00	4.60	0.57n.s.	0.29n.s.
SD	2.67	4.99		

Note. n.s. = not significant.

Table 5.13 shows that on average L2 participants produced a smaller "largest" cluster than their L1 counterparts. But the difference was too slim to produce a meaningful difference ($t(58) = 0.32$, n.s.). Thus, this type of cluster did not create a substantial impact on the overall results at all, producing a negligible effect in generating L1 and L2 differences in sorting behaviour.

Table 5.13. Mean largest cluster participants made

	NS ($n = 30$)	NNS ($n = 30$)	t-value	F-value
Mean	10.83	10.47	0.32n.s.	0.50n.s.
SD	5.13	3.63		

Note. n.s. = not significant.

Table 5.14 shows that NS produced on average a slightly smaller "largest" cluster with the sorting task using 0.5K words than with the task using 1K randomly selected verbs. In contrast, NNS produced on average a slightly smaller "largest" cluster with the sorting task using 1K verbs (Table 5.15). However, as the t-values show, both of these differences were small and not statistically significant. The revised sorting task was similar to the previous one in that both tasks produced very similar average largest cluster sizes in the case of both the NS and NNS groups.

Table 5.14. Mean largest cluster NS made: 0.5K JACET words vs. 1K JACET verbs

	0.5K ($n = 28$)	1K ($n = 30$)	t-value	F-value
Mean	9.82	10.83	0.85n.s.	0.55n.s.
SD	3.79	5.13		

Note. n.s. = not significant.

Table 5.15. Mean largest cluster NNS made: 0.5K JACET words vs. 1K JACET verbs

	0.5K ($n = 28$)	1K ($n = 30$)	t-value	F-value
Mean	10.75	10.47	0.27n.s.	1.46n.s.
SD	4.38	3.63		

Note. n.s. = not significant.

In sum, the present revised sorting task failed to produce distinct L1 and L2 differences regarding cluster number, size and variability. However, it should be noted that certain aspects of the present sorting task results analysed in this section produced the exact same tendencies as did the previous sorting task reported in Chapter 4. That is, regarding both sorting tasks, on average NS tended to produce a larger number of clusters and a smaller number of words per cluster than NNS did.

5.3.3 Native-like links in cluster structures

To answer the third research question, which is concerned with native-like links in cluster structures, I elicited pairs of native-like links that half or more of L1 participants produced in the sorting task results. For this purpose, I first constructed co-occurrence matrices showing the number of times that each individual word was associated with other words for both NS and NNS groups (See Appendices 5.1a and 5.1b). Then, as baseline data, I identified pairs of "native-like" links in Appendix 5.1a, the co-occurrence matrix of the NS group. The results of the analysis are shown in Table 5.16. The table indicates that there are 41 pairs of this category in the NS results (figures in parentheses indicate the number of participants who produced the word pair):

Table 5.16. "Native-like" links produced by NS identified in the sorting task results

accept-agree (15), *argue-discuss* (18), *ask-discuss* (17), *avoid-refuse* (15), *believe-consider* (15), *believe-imagine* (18), *build-create* (18), *build-grow* (17), *build-improve* (15), *buy-sell* (21), *buy-spend* (25), *choose-consider* (15), *choose- decide* (27), *consider-decide* (16), *describe-discuss* (16), *describe-explain* (26), *describe-speak* (15), *describe-suggest* (16), *describe-teach* (19), *describe-tell* (21), *discover-improve* (15), *discuss-explain* (15), *discuss-speak* (15), *discuss- suggest* (18), *enter-meet* (15), *explain-introduce* (15), *explain-speak* (15), *explain-suggest* (16), *explain-teach* (22), *explain-tell* (23), *get-receive* (18), *give-receive* (17), *grow-improve* (19), *hear-listen* (26), *hear-speak* (17), *join-meet* (24), *learn-read* (16), *listen-speak* (17), *read-write* (25), *sell-spend* (21), *speak-tell* (17)

Based on the baseline data identified in Table 5.16, I located the pairs in individual participant's results for both groups. Then I made a matrix that tabulated the presence or absence of the pairs of each participant for each group. In the matrix for each group, the presence of an identified pair was labelled as "1" and the absence as "0" in the cell concerned. See Appendices 5.2a and 5.2b for the relevant tabulations. An unmatched *t*-test was run on the two sets of data to search for the mean difference in native-like links between the groups. The results are shown in Table 5.17.

Table 5.17. Mean number of native-like links

	NS ($n = 30$)	NNS ($n = 30$)	*t*-value	*F*-value
Mean	25.03	19.60	3.86**	1.40n.s.
SD	4.97	5.89		

Note. ** $p < 0.01$.

Table 5.17 shows that on average L1 participants produced 25.03 native-like links while L2 participants produced 19.60 links of this type. An unmatched *t*-test revealed that there was a statistically significant difference between the two means ($t(58) = 3.86$, $p < 0.01$). Thus, the NS group produced substantially more native-like links in the present sorting task than the NNS group did. This is firm evidence for a qualitative difference of cluster structures of L1 and L2 mental lexicons, and supports previous word association test-based studies which have reported such an organisational dissimilarity (e.g., Meara, 1983; Meara & Schur, 2002; Postman & Keppel, 1970; Riegel & Zivian, 1972; Szalay & Deese, 1978; Wilks & Meara, 2002).

5.4 Discussion

The results reported in section 5.3 showed that the revised sorting task generally failed to produce the hypothesised differences in L1 and L2 lexical organisations. This was particularly true with cluster number, size and variability. Thus, the main undertaking of this section is to examine the factors behind this. In this regard, I will first discuss the use of high frequency verbs in the revised sorting task and then the directions for completing the sorting task.

The revised sorting task succeeded in finding native-like links that L1 participants made in distinctively larger number than their L2 counterparts. Thus, it was demonstrated that the revised sorting task had the sensitivity to tap into L1 and L2 differences in lexical organisation. However, the production of native-like links (41 links) was limited in number considering the 1,225 potential links that could have been made. Accordingly, the third and last issue to be discussed in this section is concerned with this relationship between native-like links and overall lexical organisation.

5.4.1 Intrinsic complexities of verbs and the revised sorting task

Ronald (2006) overviewed studies of the relationships between word class and vocabulary acquisition and referred to a "connectionist perspective" (MacWhinney, 1997). Ronald stated that "the number or strength of semantic, collocational, associative or other links which a word may have within the mental lexicon may, typically, be different depending on whether the word is a noun, a verb, or an adjective" (p.177). It was predicted that the present most high frequency 1000 verbs taken from the JACET 8000 would produce distinct L1 and L2 differences in various aspects of the sorting task results. I decided on verbs among word classes, taking into account the widely-held view of previous studies which have reported that verbs are usually the most difficult words for L2 learners. As referred to earlier, Lennon (1996) reported that simple high frequency verbs are actually the most difficult ones for L2 learners because "they [advanced-level German learners of English] may have a broad outline of verb meaning, but that their lexical knowledge is hazy concerning polysemy, contextual and collocational restrictions, phrasal verb combinations, [and] grammatical environment" (p. 35). Thus, high frequency verbs are one of the most difficult types of words for L2 learners to master.

Related to the difficulties NNS have in using verbs properly versus other types of words, Källkvist (1998, 1999) reviewed studies on the learnability of nouns and verbs (as well as words from other parts of speech) in L2 learning. She concluded that "verbs have more complex semantics and vary more cross-linguistically than nouns" (p.150), whereas nouns are usually learned predominantly in the early stages of L2 learners. Furthermore, as Gentner (1981, 1982) indicates, verbs tend to have greater breadth of meaning than other word classes. Gentner reported that the 20 most frequently used verbs in English have an average of 12.4 meanings each while the 20 most frequently used nouns have an average of 7.3 meanings each. It also should be noted that verbs for the most part are semantically the most complex word class (Miller and Fellbaum, 1991; Read, 2004). The intrinsic semantic complexities of verbs were also hypothesised to be difficult to overcome for advanced-level Japanese learners of English, who were the L2 participants in the present experiment. The sorting task results should have revealed evident differences reflecting discrepancies between the English language development and the mental lexicon organisations of NS and NNS.

There seem to be two reasons why the high frequency verbs failed to produce evident L1 and L2 differences in lexical organisation. First, there is a high possibility that the differences in aspects of lexical knowledge of high frequency verbs between the NS and NNS groups had less impact on sorting task behaviour than had been predicted. As Källkvist (1998, 1999) and Lennon (1996) reported, native speakers have richer (i.e., deeper) lexical knowledge of the tested verbs than non-native speakers do. If participants in an experiment were given a well-tuned test to tap into vocabulary depth, the results would likely reveal distinctive differences between the NS and NNS groups. However, the present sorting task was not developed in order to examine L2 learners' lexical knowledge, but rather to investigate their lexical organisation. The task asked participants to sort verbs into clusters they thought were related to each other semantically. Thus, besides the polysemous nature of verbs in task completion, the task as a whole seems to have not required participants to access the aspects of verb complexities identified above. High frequency verbs used in the revised sorting task addressing L1 and L2 lexical organisations did not produce the predicted differences, in spite of the findings of previous studies which had reported verb difficulties for L2 learners over other word classes. Of course, we should not be hasty in making a claim that word class effects on sorting task results are negligible at this stage of the project. To do so, other word classes need to be tested in future sorting tasks.

Another reason why I believe I need to conduct other sorting tasks using words selected from other word classes is derived from the comments participants made. Participants of both L1 and L2 groups in the present experiment commented after they had completed the task that the task was not easy but fun, and thought over hard to decide which meaning of a verb they should decide on. They stated that this was because verbs have different meanings in accordance with what types of words follow or come before them. This type of comment indicates that there actually existed verb-specific features that affected the participants when carrying out the task. Although the task required them to solely work out the semantic relationships among the tested verbs, participants inevitably activated their syntagmatic and collocational knowledge of the verbs during the task completion process. Thus, strictly speaking, the intrinsic complexities of the verbs affected participants' sorting behaviour, although the effect was not noticeable in the results. Considering this fact that participants activated aspects of lexical knowledge that are unique to verbs, it is predicted that non-verb words (e.g., nouns and adjectives) having less complexities in syntactic constraints would produce different results in sorting task experiments.

Second, the selection procedure for tested words and the selected tested words might have been problematic and thus hindered the task from revealing distinctive L1 and L2 differences. More specifically, selecting tested words randomly from a word list might not be a valid method for choosing words for a psycholinguistic experiment such as a sorting task. Although both the previous and present sorting tasks succeeded in identifying links in lexical organisation that L1 participants produced substantially more than their L2 counterparts, the number was extremely limited. I argue that this can be accounted for to some extent by the fact that the tested words were selected randomly without explicit relatedness. If they had been selected in a more consistent way so that participants could identify their underlying semantic relatedness, the results might have produced more distinct differences in lexical organisation between the NS and NNS groups. For future studies, it is requisite for me to devise a procedure for selecting words embracing this nature of relatedness more closely than the tested words I have used in the sorting tasks thus far. In this regard, one plausible solution is to select words out of a well-written passage that has cohesion, defined by Jaworska (1998) as "the linguistic marking of the links between a sequence of grammatically distinct sentences that make these sentences hang together" (p. 55). A text having cohesive relationships entails both grammatical and lexical cohesions. Accordingly, if tested words for sorting tasks were chosen out of a passage that has the lexical cohesion of "the cohesive effect achieved by the selection of vocabulary" (Halliday & Hasan, 1976, p. 274), they might produce more clear-cut L1 and L2 differences in the sorting task results. This is an area that needs to be revised in the word selection procedure in future sorting tasks.

5.4.2 Lexical and conceptual knowledge and task completion time

The limited results in revealing L1 and L2 differences in lexical organisation might also have been attributed to the task direction stating "You will have 20 minutes. You should have time to think it over and change it at the end before you stop." Thus, within the maximum 20 minutes, participants activated both lexical and conceptual knowledge in their mental lexicons. The activation appears to have played a different role in task completion depending on whether the participants were native speakers of English or not. The generous 20 minute time limit affected L2 participants more than their L1 counterparts. Table 5.18 shows the number of participants who took the maximum 20 minutes to complete the previous sorting task as well as the present one. The table reveals that only one L1 participant in each of the sorting tasks took the maximum 20 minutes to complete the task. Meanwhile, five (17.9%) L2 participants took the full 20 minutes in the previous sorting task and 16 (53.3%) did in the present sorting task. These results show that the 20 minute maximum time limit influenced L2 participants more than it did L1 participants. I argue that this is problematic, considering that the aim of both of the sorting tasks was to examine L1 and L2 differences in lexical organisation.

Table 5.18. Number (%) of participants who took the maximum 20 minutes to complete the sorting task

	0.5K (n = 28)	1K (n = 30)
NS	1 (3.6%)	1 (3.3%)
NSS	5 (17.9%)	16 (53.3%)

Because of the generous time limit, L2 participants in particular appeared to access their conceptual knowledge and L1 lexical knowledge while carrying out the task that intended to tap into their L2 lexical knowledge. In the end, L2 participants took more time to complete the sorting tasks than L1 participants did. That is, L2 participants seem to have activated L1-L2 lexical and conceptual links in the process of completing the sorting tasks. This was surmised from observing an L2 participant who was asked to think aloud while carrying out the 1K verb sorting task. Besides the main groups of the participants, I asked other participants (one native speaker and one advanced Japanese speaker of English) to think aloud. The Japanese participant often muttered in Japanese to elaborate the semantic relationships between the tested words. He thought over the facets of meanings of the words, the interconnections between them, and then sorted them into clusters. On the other hand, a native English speaker counterpart did the task while muttering (thinking aloud) solely in English. In the process of carrying out such a psycholinguistic task as the sorting task, it would be natural for L2 speakers, even if they are at an advanced level of the language, to use their first language from time to time. L2 participants in the sorting task made use of lexical and conceptual links between their L1 and L2 to access and process meanings. In the end, L2 participants took more time to complete the task than L1 participants did.

The fact that L2 participants took more time to complete the task might also have been due to the possibility that L2 conceptual links to concepts (i.e., meanings) are weaker than the L1 conceptual links to them. Kroll and Tokowicz (2001) proposed the Revised Hierarchical Model to explain this phenomenon. The model, which is shown in Figure 5.1, depicts three distinctive features of a bilingual lexicon: (a) at the lexical level, connections from L2 to L1 are stronger than the connections from L1 to L2; (b) the size of L1 is bigger than that of L2 and (c) at the conceptual level, the connections of L1 words are stronger than those of L2 words.

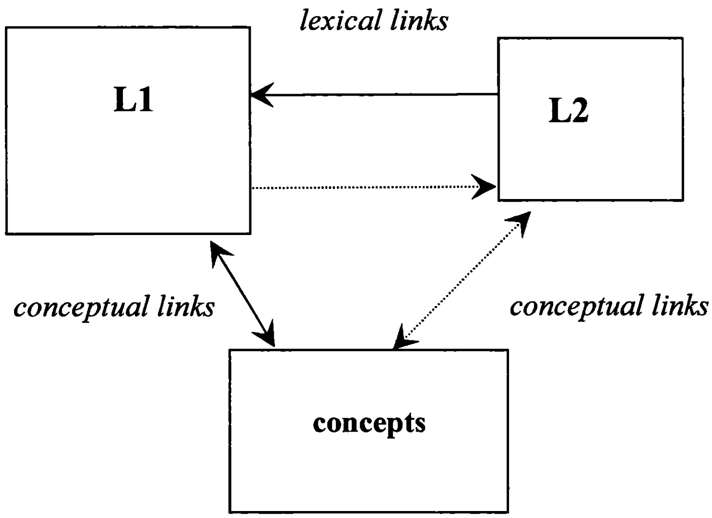

Figure 5.1. Revised Hierarchical Model (adapted from Kroll & Stewart, 1994). Words in each language (L1 and L2) are interconnected via lexical-level links and conceptual links. The lexical-level links are stronger from L2 to L1 (solid line) than from L1 to L2 (dashed line) but the conceptual links are stronger for L1 (solid line) than for L2 (dashed line) (Kroll & Tokowicz, 2001, p.51).

Kroll and Tokowicz (2001) suggested that "the lexical-level connections established during early stages of acquisition may still continue to function under some circumstances once individuals become fluent bilinguals" (p. 52). That is, "under some circumstances", the meanings of L2 words are mediated through the L1 even if the L2 speakers in question have reached an advanced level. The nature of the sorting task in the present experiment might have been one of those "circumstances" where L2 participants had to activate their conceptual knowledge to reach a satisfactory decision. In such a situation, even advanced-level L2 speakers often access their L1 to better understand the meanings of L2 words and the interconnections between them. This might in part account for the fact that L2 participants took more time to complete the sorting task than their L1 counterparts.

Given these together, it is natural that a significantly larger number of L2 participants took the maximum allotted time of 20 minutes to complete the task. Perhaps it would be almost impossible for L2 participants to complete such a task as a sorting task by only accessing their L2 lexical knowledge and not their L1 lexical knowledge and conceptual knowledge. However, it was found that the instructions, which permitted participants to take the maximum 20 minutes to complete the task, affected L2 participants in particular. In future experiments, the directions should be revised so that the sorting task is able to tap more into the L2 lexical knowledge of participants than their L1 lexical knowledge and conceptual knowledge.

5.4.3 Native-like links and L1/L2 differences in lexical organisation

The present sorting task confirmed that L1 lexical organisation is different from its L2 counterpart as shown by the results of an unmatched *t*-test for native-like links. Examination of the co-occurrence matrices revealed that there were 41 "native-like" links (e.g., *accept-agree*, *describe-suggest*, *join-meet*) where a distinctive between-group difference was found. As stated earlier, however, it should be noted that this L1 and L2 difference was extremely limited in number, considering the overall structure of the mental lexicon. Calculations for the number (%) of blank cells (paired links that no participant produced), the number (%) of filled-in cells (paired links that at least one participant produced) and the ratio of the 41 "native-like" links against the overall links identified in the co-occurrence matrices are shown in Table 5.19.

Table 5.19. Number (%) of blank cells (linked pairs), filled-in cells and native-like cells (links) in the co-occurrence matrices (cell number total = 1,225)

	No. (%) of blank cells	No. (%) of filled-in cells	No. (%) of native-like cells (links)
NS ($n = 30$)	196 (16.0%)	1029 (84.0%)	41 (3.9%)
NNS ($n = 30$)	256 (20.9%)	969 (79.1%)	41 (4.2%)

Note. The percentages of native-like cells (links) were calculated by dividing the number of native-like cells by the number of filled-in cells.

Table 5.19 reveals that only 3.9% of the cells (paired links) in the matrix were native-like links for L1 participants and only 4.2% for L2 participants. As for a possible reason for these low percentages, it seems to be attributed to the large individual differences found in cluster number, size and variability between participants that were reported in section 5.3. In particular, this is concerned with a lack of homogeneity in the results of the L1 participants. The mean number of clusters (which excludes single, isolated words) NS produced was 9.93 and the *SD* was 3.05. Meanwhile, the mean number of words per cluster (which excludes single, isolated words) was 5.39 and the *SD* was 2.75. (See Tables 5.4 and 5.7 for details.) The large values of the *SD*s imply that the sorting task results of the NS group tabulated in the co-occurrence matrix were not consistently homogeneous. Thus, L1 participants failed to produce a significant number of native-like links. As Fitzpatrick (2007, 2008) reported, native speakers are not always homogeneous in their response behaviour in word association tests. She cautions postulating "native-like" responses in psycholinguistic data elicitation experiments. The results of the present sorting task support her claim. Although a hasty conclusion should not be made, in future studies an examination of L1 and L2 cluster structures in the results of a sorting task should be made by another approach besides native-like links. This issue will be addressed in experiments using further revised sorting tasks in the following chapters.

Given the limited number of L1 and L2 differences identified above, it is also possible that differences in lexical organisation might be subtler than hypothesised. This subtlety was reported by Wilks (1999). She predicted that the difference between the overall structures of L1 and L2 lexical networks may be more subtle than expected, while differences in associative patterns might exist that affect the way words in the lexicon are interconnected. Therefore, regarding this issue of subtlety as well, it is necessary to conduct other experiments after making revisions to the present sorting task while at the same time correcting the weaknesses identified above.

It is highly plausible that L1 and L2 lexical networks are different from each other regarding how the links between words are disposed. That is, what is crucial for L1 and L2 differences in cluster structures has to do with the number, size, and variability of clusters as well as their arrangement. Wilks and Meara (2002) claimed that a simple comparison of the average number of connections between L1 and L2 lexical networks may be more misleading than generally assumed. They predicted that "two networks with the same density could in fact be quite differently arranged in terms of how the connections between points are disposed" (Wilks & Meara, 2002, p. 319). In future studies, this issue of how the two lexical organisations are structured and how different they are from each other will also be addressed.

5.5 Conclusion

Contrary to my prediction, the present revised sorting task as a whole failed to tap into any distinct L1 and L2 differences in lexical organisation. Although simple high frequency verbs had been hypothesised to be the most difficult even for L2 learners at an advanced-proficiency level and to bring about some distinctive differences, the sorting task using them did not produce any of the assumed evident group differences. However, a closer examination of the low task sensitivity revealed that the lack of any distinctive differences should not be attributed to the verbs used in the sorting task but instead to two other factors in the task itself. One was concerned with the fact that the tested words were selected randomly from the JACET 8000 and the semantic relatedness between them was not easy for participants to ascertain and sort into clusters. Another problem was attributed to the task directions that permitted participants to take up to the maximum 20 minutes to complete it. This generous time limit affected L2 participants more by allowing them to activate their conceptual knowledge and L1 lexical knowledge as well as their L2 lexical knowledge. Thus, in future experiments, to improve the task sensitivity in examining lexical organisation, I will revise the sorting task so that participants will be able to find the semantic relatedness between words more easily and carry out the task while activating their L2 lexical knowledge more than their conceptual knowledge and L1 lexical knowledge.

The investigation of native-like links in the task results revealed that L1 participants were actually different from their L2 participants regarding their lexical organisation. However, the detected difference was small in size accounting for only 3.9% of the total links NS produced and only 4.2% of the total links NNS produced. The difference might be subtle as the present results seem to indicate, but it might have resulted from the lack of task sensitivity identified above. It should be noted that in future studies I also need to address how the links between lexical items in the mental lexicon are disposed. That is another important facet of lexical organisation that should be further examined.

In Chapter 6, I will develop another sorting task that should boost its ability to tap into L1 and L2 differences in cluster structures of lexical organisation by resolving the identified weaknesses in the present sorting task. I will then further address whether there are substantial differences in aspects of lexical organisation between NS and NNS.

Chapter 6: Boosted task sensitivity and examination of L1 and L2 differences in sorting behaviour

6.1 Introduction

In Chapter 5, using 50 verbs randomly selected from 1K (the most frequent 1000) words in the JACET 8000, it was confirmed that L1 lexical organisation is different from its L2 counterpart as shown by an unmatched *t*-test for native-like links, i.e., the links that half or more of L1 participants produced in the results. Thus, broadly speaking, the previous sorting task proved to be a promising method to tap into the differences in cluster structures of L1 and L2 mental lexicons.

Meanwhile, the sorting task failed to reveal any differences between the overall structures of L1 and L2 lexical organisations. The NS group produced a larger number of clusters, a smaller cluster size and they were less varied than their NNS counterpart as was the case with Meara and Schur's (2002) research, but many of the differences were not substantial ones. Moreover, the 41 native-like links detected accounted for only a very small portion of the overall structures. In theory, with 50 words for a sorting task, there could have been 1,225 potential pairs of linked words in the results, but the cut-off point approach analysis accounted for only 41 of them. In other words, 96.7% of the data was left untouched and not reflected in the analysis. As Wilks (1999) predicted and the results in the previous sorting task showed, the differences of the overall structures in L1 and L2 lexical organisations might be more subtle than expected. Advanced-level L2 speakers, who were participants in the experiment in the previous sorting task, might have developed almost the same L2 lexical organisation as their L1 counterparts at least regarding the high frequency verbs. The results gained so far suggest this to be the case.

However, it is still too early to make a conclusion. This is because the low sensitivity in the previous sorting task seems to have been caused by two prohibitive factors. One was a problem in the word selection procedure for the task. The 50 verbs were chosen randomly from the JACET 8000, and there was a lack in semantic relatedness between them. Eventually participants in the experiment found it difficult to ascertain the relationships between words and reliably sort the words into clusters. Put in another way, by using words of close semantic relatedness in a sorting task, participants might make less single, isolated words and a bigger largest cluster of linked words. Therefore, the words for a revised sorting task should be chosen so that participants can find their semantic relatedness easily. Another problem lies in the task direction of "You will have 20 minutes. You should have time to think it over and change it at the end before you stop." This affected particularly L2 participants because they activated their conceptual knowledge and L1 lexical knowledge as well as their L2 lexical knowledge. This is against the purpose of the present research project which aims to reveal L1 and L2 differences in lexical organisation. For a revised sorting task, the directions should be devised in a way that participants can complete the task by instinctively activating their lexical knowledge more than their conceptual knowledge.

In this chapter, three multivariate analyses (i.e., pathfinder analysis, Multidimensional Scaling (MDS) and cluster analysis) were run on the sorting task results and compared with each other. These three analyses are the most widely-used ones to "measure, explain, and predict the degree of relationship" (Hair, Black, Babin, Anderson & Tatham, 2005, p. 4) among tested variables. Thus, analyses can be validly made on the data elicited by a psycholinguistic experiment to reveal the underlying organisational structure. (See Chapter 2 for a comparison of multivariate analyses). The purpose of making the comparison is to examine which is the most reliable analysis in probing into L1 and L2 differences in sorting task results and to decide on one multivariate analysis for the present and future experiments in this research project. This is also motivated by the decision that at this stage of the research project it is appropriate to address the issue of the arrangement of L1 and L2 lexical organisation by means of multivariate analyses.

This chapter addresses the same research questions on task completion time (research question 1a) and cluster number, size and variability (research question 2a) as in the previous chapter, while newly addressing whether the revised sorting task has increased the sensitivity of the testing method compared to the earlier methodology (research questions 1b and 2b). It also addresses a comparison of three types of multivariate analyses (research question 3). These points are summarised below:

1a. Do L1 participants take less time to complete a sorting task than L2 participants?
1b. Does the present revised sorting task take participants less time to complete than the previous one?
2a. Do L1 participants make a larger number of clusters and fewer words per cluster than their L2 counterparts do? Is L1 lexical organisation less varied than its L2 counterpart?
2b. Regarding the variables shown in 2a, as well as the single, isolated words and the largest cluster made, are there distinctive differences between the results produced by the previous sorting task and the present one?
3. Which multivariate analysis (i.e., pathfinder analysis, MDS or cluster analysis) is the most reliable in examining L1 and L2 differences in cluster structures?

6.2 Method

6.2.1 Participants

In this study, there were two participant groups. The first was comprised of 30 adult, native speakers of English (NS). They were teachers of English in Kumamoto, visiting scholars and students studying at Kumamoto University. The second consisted of 30 adult, advanced-level Japanese speakers of English (NNS). They were either English teachers at the college level or persons having a high competence of English as judged by a TOEFL score of 213 or more on the computer-based version or a score of 550 or more on the paper-based version or a TOEIC score of 730 or more that had been taken within the last two years.

6.2.2 Data collection

Fifty verbs were used in the present sorting task. To make it easier for participants to sort into groups according to word meaning, one chapter from a book was chosen. I first chose a chapter from Robert Stevenson's (1883), *Treasure Island*, "Chapter 32. The Treasure-hunt — The Voice Among the Trees". It was selected because the chapter has verbs suited for the purpose of the present experiment. The 50 verbs from among the verbs contained in the chapter, which are among the first, 1000 high frequency words in the JACET 8000, were randomly selected: *to add, to ask, to beat, to begin, to believe, to break, to bring, to come, to cry, to describe, to die, to face, to fall, to feel, to fight, to find, to get, to go, to have, to hear, to help, to hold, to increase, to keep, to leave, to listen, to look, to make, to mind, to remain, to reply, to rest, to return, to rise, to run, to say, to see, to show, to sing, to sit, to speak, to stare, to start, to stop, to strike, to struggle, to take, to tell, to think, to walk.*

After piloting, the task was administered to each of the participants individually. Each pile of 50 cards, on which the individual English words were printed, was shuffled and bound with a rubber band and put into an envelope for each participant. The order of the cards was fixed so that participants would read them in the same order. Participants were invited to sort them into groups of words that they thought would go together according to meaning. In the case that participants found words that they thought didn't seem to fit into any of the word groups they made, they could leave these as single, isolated words. It didn't matter how many groups they made. Participants were directed to do the task as quickly as possible so that they would complete the task while intuitively activating their L2 lexical knowledge more than their L1 lexical knowledge and conceptual knowledge. The directions were written in English for native speakers of English, and in Japanese for Japanese participants (see Appendix 6.1 for the English version). Participants read through them, asked questions if they had any, and then worked on the task. They got a small present when they finished. As in the previous experiments, the time that participants needed to read the cards and to sort them were not separately recorded, and the total was regarded as the time they took to complete the sorting task.

6.2.3 Data analysis

To answer research questions 1a and 2a, the same analyses were made as in Chapter 5. Accordingly, unmatched *t*-tests were run to examine L1 and L2 differences in the mean task completion time, mean number of clusters and mean number of words per cluster the NS and NNS groups produced. The mean numbers of single, isolated words and the mean largest cluster that participants made were also compared to each other. In addition, unmatched *t*-tests were run on the data produced by the previous sorting task and the present one. This was done to examine whether the revisions made to the present sorting task helped boost the task sensitivity to tap into L1 and L2 differences in lexical organisation (research questions 1b and 2b).

To answer research question 3, three multivariate analyses (i.e., pathfinder analysis, MDS and cluster analysis) to probe into the organisational differences in L1 and L2 lexical structures were adopted.

Pathfinder analysis derives a pathfinder network from proximities for pairs of entities. Proximities are gained from similarities, correlations, distances and conditional probabilities of the relationships among entities. In the data collected by a sorting task, entities represent tested words, and entities correspond to the nodes of the generated associative network. In addition, the links in the network are determined by the patterns of proximities. Thus, explicit links represent weighted paths between nodes. The weights in the present sorting task data are frequency counts for each pair of links in the co-occurrence matrix (i.e., the number of participants who produced the paired link in question). Therefore, the higher a frequency is, the higher its similarity degree is. The computed links are drawn in order to exclude all but the single highest counts for word pairs, which eventually show a lexical network to only the most salient relationships. (See Sánchez (2004) for an example of an application of this technique to a psycholinguistic experiment of L2 vocabulary instructional effects.)

Multidimensional Scaling (MDS) is a mathematical technique by which the positions of data points are computed and decided so that the proximities among data can be expressed in the best way. The computed results are drawn in the form of a two or three dimensional Euclidean representation. The advantage of MDS is that it gives a visual representation where entities that have a high degree of proximity are placed close to each other and those that do not are placed far away from each other. Therefore, in the case of the present experiment, how the tested 50 verbs are placed in a two or three dimensional representation is shown by applying MDS computation to the collected data. See Rapoport and Fillenbaum (1972) and Routh

(1994), which were reviewed in Chapter 2, for examples of applying the MDS technique to tap into L1 cognitive structures. Following the method of these previous studies, a two dimensional representation for the present sorting task results is constructed.

Cluster analysis is a mathematical method to group "individuals or objects into clusters so that objects in the same cluster are more similar to one another than they are to objects in other clusters" (Hair et al., 2006, p. 555). Cluster analysis starts by merging two clusters having the highest degree of similarity (proximity) in the data matrix and finishes by the whole data being merged into a single cluster. The completed calculation will be expressed by drawing a dendrogram (tree diagram) for it. See Miller (1969), Rapoport and Fillenbaum (1972) and Routh (1994) for their use of cluster analysis in examining memory structures.

A comparison of the results obtained by these multivariate analyses will be conducted, addressing which is the most reliable analysis for examining L1 and L2 differences in the results of the present sorting task.

6.3 Results

6.3.1 Time taken to complete task

To answer research questions 1a and 1b addressing L1 and L2 differences in task completion time, relevant information is presented in Tables 6.1 to 6.3. Table 6.1 shows the means and *SD*s of the amount of time that the NS and NNS groups took to complete the sorting task using the 50 high frequency verbs selected from *Treasure Island.* On average, NNS took 1.61 minutes more to finish the task than NS did. However, an unmatched *t*-test revealed that there was no statistically significant difference between the two means ($t(58) = 1.73$, n.s.). This was distinctively different from the results the previous sorting task produced, where NNS took 3.04 minutes more to complete it than NS did and the difference was statistically significant. There was only one L2 participant who took more than 20 minutes to complete the task, whereas in the previous experiment 16 L2 participants had taken 20 minutes, which was the maximum time allowed in the previous sorting task. Thus, it was found that using verbs which have a high cohesion under the direction to sort them as quickly as possible led both NS and NNS groups to complete the task in a similar way.

Table 6.1. Time to complete 1K *Treasure Island* verbs sorting task

	NS ($n = 30$)	NNS ($n = 30$)	*t*-value	*F*-value
Mean	7.94	9.55	1.73n.s.	1.40n.s.
SD	3.28	3.89		

Note. n.s. = not significant.

Addressing the question of whether revisions to the present sorting task had an effect on task completion time, Table 6.2 shows that L1 participants took 6.82 minutes less to complete the revised sorting task using 1K *Treasure Island* verbs than the previous one using 1K randomly selected verbs from the JACET 8000. An unmatched *t*-test revealed that there was a statistically significant difference between the two means ($t(58) = 7.32, p < 0.01$). Thus, the revisions made to the present sorting task led L1 participants to complete the task far more quickly than the previous sorting task, suggesting that they completed the task more instinctively.

Table 6.2. NS's time to complete the sorting task: 1K JACET verbs vs. 1K *Treasure Island* verbs

	JACET (n = 30)	*Treasure* (n = 30)	*t*-value	*F*-value
Mean	14.76	7.94	7.32**	1.42n.s.
SD	3.91	3.28		

Note. JACET = the 50 verbs randomly selected from verbs in the most frequent 1000 words in the JACET 8000; *Treasure* = the 50 verbs randomly selected from verbs in *Treasure Island* that are contained in the most frequent 1000 words in the JACET 8000. The same notations will be used hereafter in this chapter. n.s. = not significant; $^{**}p < 0.01$.

Table 6.3 shows that L2 participants took 8.25 minutes less to complete the revised sorting task. An unmatched *t*-test revealed that there was a substantial difference between the two means ($t(58) = 8.83, p < 0.01$). This shows that the revisions to the present sorting task were effective in having L2 participants complete the task more quickly, as was the case with L1 participants. It should be noted that L2 participants' shortening of task completion time of the present sorting task over the previous one (8.25 minutes) outweighed that of L1 participants' (6.82 minutes). It should be suggested that, with the revised sorting task, L2 participants sorted the verbs into clusters while instinctively accessing their L2 lexical knowledge more than their L1 lexical knowledge and conceptual knowledge.

Table 6.3. NNS's time to complete the sorting task: 1K JACET verbs vs. 1K *Treasure Island* verbs

	JACET (n = 30)	*Treasure* (n = 30)	*t*-value	*F*-value
Mean	17.80	9.55	8.83**	0.73n.s.
SD	3.32	3.89		

Note. n.s. = not significant; $^{**}p < 0.01$.

Given these results as a whole, the revisions to the sorting task can be said to be effective. The present sorting task, which revised the word selection procedure and the direction in task completion time, produced the desired result that participants would work on the task while activating their lexical knowledge more than their conceptual knowledge.

6.3.2 Cluster number, size and variability

This section will answer research questions 2a and 2b addressing L1 and L2 differences in cluster number, size and variability. Analyses will be made by comparing the NS and NNS groups regarding the present sorting task results, followed by a comparison between the previous sorting task and the present revised one. For this purpose, four sets of tables (mean number of clusters, mean number of words per cluster, mean number of single, isolated words, and the mean largest cluster participants made) will be shown. Each of the tables has three related tables.

Table 6.4 tabulates the mean number of clusters participants made in the present sorting task, where the count excludes single, isolated words. The table reveals that on average NNS generated a slightly larger mean cluster number than NS did, but the difference was not statistically significant ($t(58) = 0.40$, n.s.). As the *SD*s show, the variance of the NS results was smaller than that of the NNS results. However, an *F*-test revealed that there was no substantial difference in variability between the two groups ($F(29) = 1.03$, n.s.). Thus, the revised sorting task failed to distinguish NS from NNS in mean cluster number and the variability that participants generated.

Table 6.4. Mean number of clusters (which excludes single, isolated words)

	NS ($n = 30$)	NNS ($n = 30$)	t-value	F-value
Mean	8.40	8.70	0.40n.s.	1.03n.s.
SD	2.87	2.91		

Note. n.s. = not significant.

Table 6.5 shows that L1 participants produced on average a smaller number of clusters with the sorting task using 1K *Treasure Island* verbs than they did with the task using 1K randomly selected verbs. An unmatched t-test revealed that there was a statistically significant difference between the two means (($t(58) = 2.01$, $p < 0.05$). It should be noted that the *SD* of the revised sorting task results was smaller than that of the previous one, although the difference between the two sorting tasks were not substantial. These results suggest that with the revised sorting task L1 participants more instinctively identified which words had semantic relatedness and which ones did not, and the variability among participants became smaller.

Table 6.5. Mean number of clusters for L1 participants (which excludes single, isolated words): 1K JACET verbs vs. 1K *Treasure Island* verbs

	JACET (n = 30)	*Treasure* (n = 30)	t-value	F-value
Mean	9.93	8.40	2.01*	1.13n.s.
SD	3.05	2.87		

Note. n.s. = not significant; * $p < 0.05$.

Table 6.6 shows that L2 participants similarly produced on average a smaller number of clusters with the sorting task using 1K *Treasure Island* verbs than they did with the task using 1K randomly selected verbs. However, an unmatched t-test revealed that there was no statistically significant difference between the two means (($t(58) = 0.77$, n.s.). As was the case with the NS group, the *SD* of the revised sorting task results was smaller than that of the previous one. These results suggest that with the revised sorting task L2 participants more instinctively discerned which words had semantic relatedness and which ones did not, but the effect was smaller than that of L1 participants. Furthermore, the variability among L2 participants turned out to be less as was the case with L1 participants.

Table 6.6. Mean number of clusters for L2 participants (which excludes single, isolated words): 1K JACET verbs vs. 1K *Treasure Island* verbs

	JACET (n = 30)	*Treasure* (n = 30)	t-value	F-value
Mean	9.33	8.70	0.77n.s.	1.37n.s.
SD	3.41	2.91		

Note. n.s. = not significant.

Table 6.7 shows the mean number of words per cluster participants in the present sorting task made, where the count excludes single, isolated words. The table reveals that on average NS generated a slightly larger mean number of words per cluster than NNS did, but the difference did not reach a statistically significant level ($t(58) = 0.25$, n.s.). As the *SD*s show, the NS's variance was smaller than that of their NNS counterparts. However, an F-test revealed that there was no substantial difference in variability between the two groups ($F(29) = 1.62$, n.s.). The present sorting task failed to distinguish NS from NNS in terms of the mean number of words per cluster and the variability that participants produced.

Table 6.7. Mean number of words per cluster (which excludes single, isolated words)

	NS (n = 30)	NNS (n = 30)	t-value	F-value
Mean	6.36	6.16	0.25n.s.	1.62n.s.
SD	2.63	3.36		

Note. n.s. = not significant.

Table 6.8 shows that L1 participants produced on average a larger number of words per cluster with the sorting task using 1K *Treasure Island* verbs than they did with the task using 1K randomly selected verbs. An unmatched t-test revealed that there was no marked difference between the two means (($t(58) = 1.39$, n.s.). The mean difference did not reach a statistically significant level, but the results suggested that L1 participants discerned the semantic relatedness between more verbs selected from a cohesive passage of *Treasure Island* than that of the randomly selected verbs.

Table 6.8. Mean number of words per cluster for L1 participants (which excludes single, isolated words): 1K JACET verbs vs. 1K *Treasure Island* verbs

	JACET ($n = 30$)	*Treasure* ($n = 30$)	t-value	F-value
Mean	5.39	6.36	1.39n.s.	1.09n.s.
SD	2.75	2.63		

Note. n.s. = not significant.

Table 6.9 shows that L2 participants produced on average a larger number of words per cluster with the sorting task using 1K *Treasure Island* verbs than they did with the task using 1K randomly selected verbs. An F-test reveals that the variances of the two data were not equal ($F(29) = 0.44, p < 0.05$). Therefore, instead of a Student's t-test for testing a statistically significant difference between the two means, a Welch's t-test was run. The results showed that the mean difference did not reach a statistically significant level ($t(58) = 0.81$, n.s.).

Table 6.9. Mean number of words per cluster for L2 participants (which excludes single, isolated words): 1K JACET verbs vs. 1K *Treasure Island* verbs

	JACET ($n = 30$)	*Treasure* ($n = 30$)	t-value	F-value
Mean	5.57	6.16	0.81n.s.	0.44*
SD	2.21	3.36		

Note. n.s. = not significant; * $p < 0.05$.

Thus, the sorting task using 1K *Treasure Island* verbs had a tendency to produce a larger number of words per cluster than that of the sorting task using 1K random words for the two groups, but the differences between the two tasks were not statistically significant ones.

Analyses of between-group and between-task comparisons regarding the mean number of single, isolated words, and the mean largest cluster participants produced are shown in Table 6.10 through Table 6.15.

Table 6.10 shows the mean number of single, isolated words participants produced in the present sorting task. The table reveals that on average NS generated a smaller mean number of single, isolated words than NNS did, but the difference did not reach a statistically significant level ($t(58) = 1.60$, n.s.). Thus, the number of single, isolated words did not have a substantial impact on the overall results and failed to display any differences between the two groups.

Table 6.10. Mean number of single, isolated words: 1K *Treasure Island* verbs

	NS ($n = 30$)	NNS ($n = 30$)	t-value	F-value
Mean	3.27	4.89	1.60n.s.	1.21n.s.
SD	3.69	4.06		

Note. n.s. = not significant.

Table 6.11 shows that NS produced on average a slightly smaller mean number of single, isolated words with the sorting task using 1K *Treasure Island* verbs than with the task using 1K randomly selected verbs. An unmatched t-test revealed that there was no statistically significant difference between the two means (($t(58) = 0.38$, n.s.). Thus, the effect of selecting words from a cohesive passage for the revised sorting task was subtle and failed to reduce the number of single, isolated words produced by the NS group.

Table 6.11. NS's number of single, isolated words: 1K JACET verbs vs. 1K *Treasure Island* verbs

	JACET ($n = 30$)	*Treasure* ($n = 30$)	t-value	F-value
Mean	3.63	3.27	0.38n.s.	1.04n.s.
SD	3.76	3.69		

Note. n.s. = not significant.

Table 6.12 shows that NNS produced on average a slightly larger mean number of single, isolated words with the sorting task using 1K *Treasure Island* verbs than with the task using 1K randomly selected verbs. However, an unmatched t-test revealed that there was no statistically significant difference between the two means (($t(58) = 0.23$, n.s.). Thus, the revisions made to the present sorting task had little effect on the mean number of single, isolated words L2 participants produced.

Table 6.12. NNS's number of single, isolated words: 1K JACET verbs vs. 1K *Treasure Island* verbs

	JACET (n = 30)	*Treasure* (n = 30)	t-value	F-value
Mean	4.60	4.89	0.23n.s.	1.51n.s.
SD	4.99	4.06		

Note. n.s. = not significant.

Table 6.13 shows that on average L2 participants produced a smaller "largest" cluster than their L1 counterparts. But the difference was too small to produce a meaningful difference ($t(58) = 0.53$, n.s.). Thus, the production of a "largest" cluster did not cause an appreciable effect regarding L1 and L2 differences in sorting behaviour.

Table 6.13. Mean largest cluster participants made

	NS (n = 30)	NNS (n = 30)	t-value	F-value
Mean	12.40	11.70	0.53n.s.	1.23n.s.
SD	4.79	5.41		

Note. n.s. = not significant.

Table 6.14 shows that NS produced on average a bigger "largest" cluster with the sorting task using 1K *Treasure Island* verbs than with the task using 1K randomly selected verbs. Similarly, NNS produced on average a bigger "largest" cluster with the sorting task using 1K *Treasure Island* verbs than with the task using 1K randomly selected verbs (Table 6.15). These results suggest that the use of verbs having closer semantic relatedness in the revised sorting task led participants in both groups to make a bigger "largest" cluster than in the previous sorting task. However, as the t-values show, both of these differences were not statistically significant.

Table 6.14. Mean largest cluster participants made for L1 participants: 1K JACET verbs vs. 1K *Treasure Island* verb

	JACET (n = 30)	*Treasure* (n = 30)	t-value	F-value
Mean	10.83	12.40	1.22n.s.	1.15n.s.
SD	5.13	4.79		

Note. n.s. = not significant.

Table 6.15. Mean largest cluster participants made for L2 participants: 1K JACET verbs vs. 1K *Treasure Island* verbs

	JACET (n = 30)	*Treasure* (n = 30)	t-value	F-value
Mean	10.47	11.70	1.04n.s.	0.45n.s.
SD	3.63	5.41		

Note. n.s. = not significant.

Given these results as a whole, contrary to the prediction, the effect of the boosted sensitivity of the revised sorting task to increase the mean "largest" cluster participants made was too slim to produce substantial differences between the previous and present sorting tasks.
In sum, the present revised sorting task as a whole failed to produce noteworthy L1 and L2 differences regarding cluster number, size and variability. The two lexical organisations might have the same degree of density (the extent to which words in the mental lexicon are linked together) at least in regard to 1K *Treasure Island* verbs. That is, the differences in the two lexical organisations appear to be more subtle than one would expect as Wilks (1999) predicted. The results failed to provide evidence for Meara and Schur (2002), who reported substantial differences in aspects of lexical organisation between NS and NNS groups.

6.3.3 Multivariate analyses of the sorting task results

To answer the third research question, which addresses which multivariate analysis is the most reliable in exploring L1 and L2 differences in cluster structures produced by the sorting task results, I constructed co-occurrence matrices showing the number of times that each individual word was associated with other words for both the NS and NNS groups (See Appendices 6.2a and 6.2b).

6.3.3.1 Pathfinder analysis

Pathfinder analysis was run on the co-occurrence matrices. In the analysis, the computer programme *Pathfinder* (Version 5.4) was employed. Following the directions of the software, two parameters were set. (1) The q-parameter constrains the number of indirect proximities examined in generating the network. The q-parameter is an integer value between 2 and n-1, inclusive, where n is the number of nodes or items. In the case of the present sorting task data, the q-parameter is 49 (50 minus 1). (2) The r-parameter defines the metric used for computing the distance of paths (e.g., the Minkowski r-metric). The r-parameter is a real number between 1 and infinity, inclusive. For the present data, the r-parameter is set as infinity, where "the path weight is the same as the maximum weight associated with any link along the path" (Dearholt & Schvaneveldt, 1990, p. 3). Both of the parameters have the effect of decreasing the number of links in the network as their values are increased. The pathfinder associative network of the NS group based on the computation is shown in Figure 6.1 and that of the NNS group is shown in Figure 6.2.

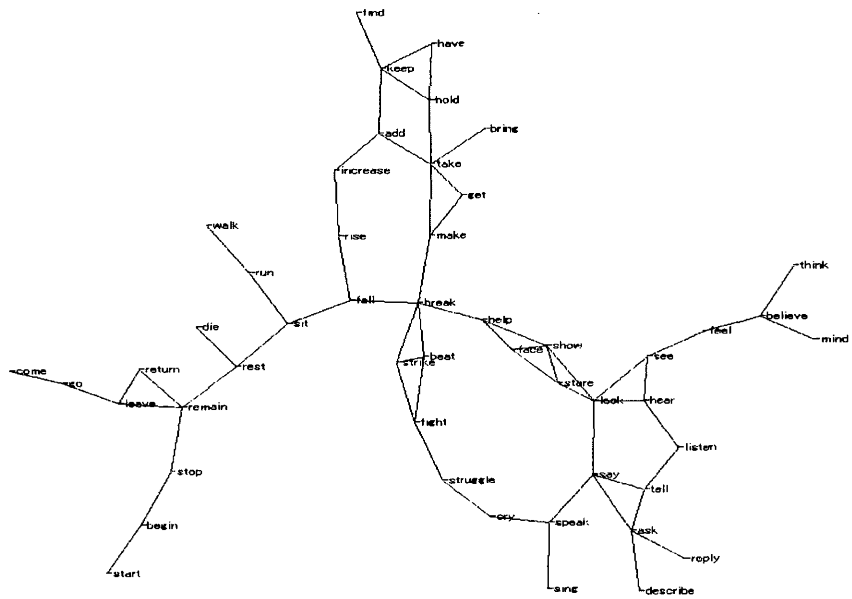

Figure 6.1. Pathfinder associative network of NS participants sorting 50 1K *Treasure Island* verbs into clusters of related meanings (n = 30)

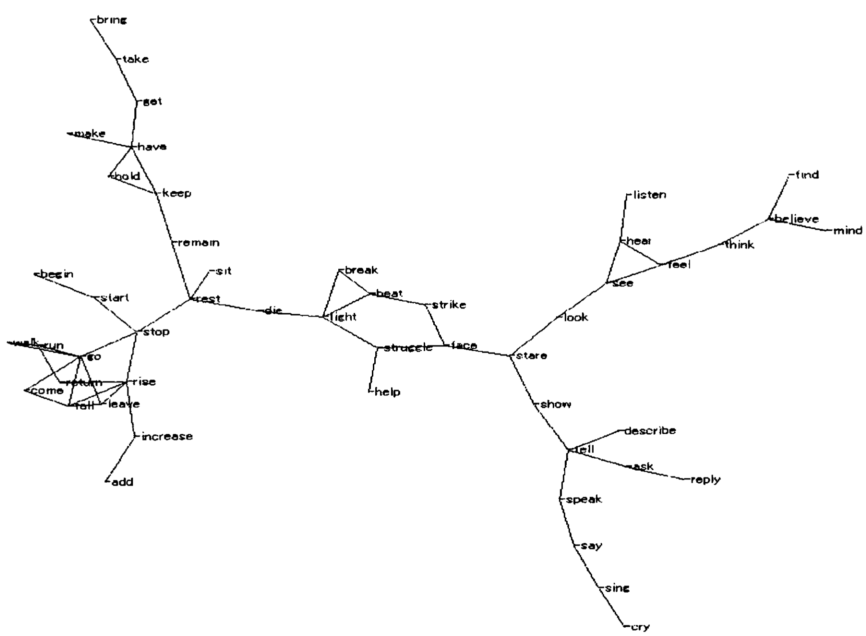

Figure 6.2. Pathfinder associative network of NNS participants sorting 50 1K *Treasure Island* verbs into clusters of related meanings (n = 30)

In the figures, the geodesic distance, relating to the shortest possible line between two nodes in a network structure, is the "length of the minimum-cost path connecting the nodes" (Chen, 2003, p. 105). Thus, in the information visualisation, the closer nodes are, the more closely related they are.

By comparing Figures 6.1 and 6.2, at first sight the pathfinder analysis seems to have succeeded in revealing the structures of and the differences in L1 and L2 lexical organisations. L1 lexical organisation was made up of seven major partitioned branches, while its L2 counterpart was composed of five major partitioned branches. The relationships between the lexical items of each branch can easily be grasped. Furthermore, in the L1 organisation, *break* and *look* were the hub nodes of the structure, and each hub node had five direct links to other node words. Meanwhile, the L2 organisation did not have such an evident hub node of the structure. Thus, it was found that the pathfinder analysis detected the features and differences of L1 and L2 lexical organisations at a local level of the sorting task results. In short, the visualisation made by the pathfinder analysis was "an effective link-reduction mechanism" (Chen, 2003, p. 106). However, on the other hand, it was not easy to grasp the overall relationships between the 50 verbs tested in the experiment and the differences between L1 and L2 cluster structures. Pathfinder analysis did not provide us with any information of how the seven major branches the NS group made were related to each other. This was also true with the five major branches the NNS group made as well.

6.3.3.2 MDS (Multidimensional Scaling)

Figures 6.3 and 6.4 represent MDS computations and their two-dimensional Euclidean representations of the NS and NNS results, respectively. Figure 6.3 shows that MDS detected three clear, closely related verb clusters in the NS sorting task results. They were:

(1) the verbs of motion of GO in the (+X, +Y) dimension (15 words): *die, remain, stop, begin, come, sit, start, go, rest, leave, return, walk, run, fall* and *rise*
(2) the verbs of motion of HAVE and FIGHT in the (+X, -Y) dimension (16 words): *bring, make, get, hold, keep, find, take, have, break, increase, help, strike, add, struggle, beat* and *fight*
(3) the verbs of SPEAK and THINK in the (-X, +Y) dimension (15 words): *speak, sing, hear, say, see, stare, reply, look, tell, think, describe, ask, feel, cry* and *listen.*

In addition, *believe, mind, face* and *show* in the (-X, -Y) dimension are separately represented, and thus cannot be regarded as a cluster.

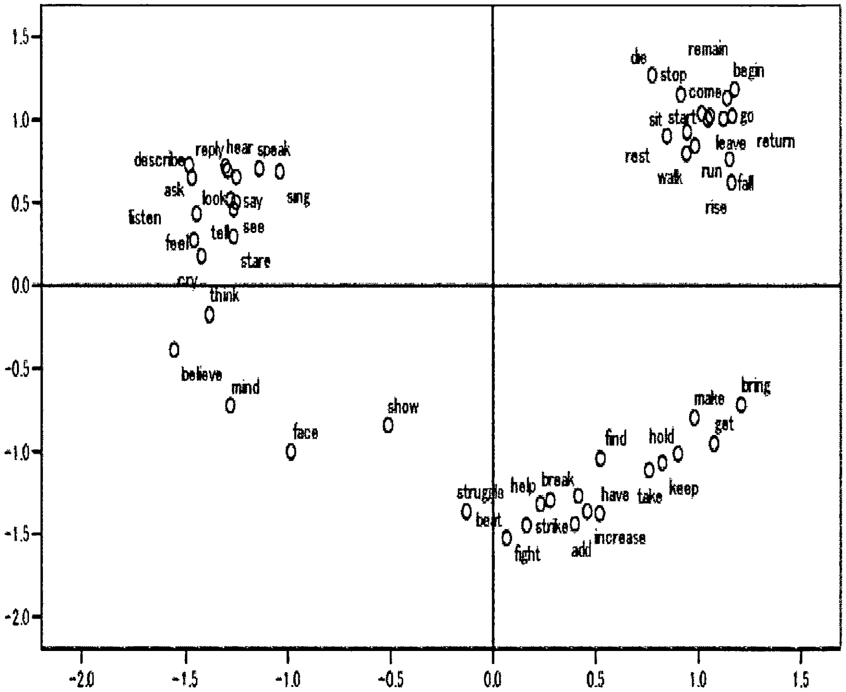

Figure 6.3. Two-dimensional Euclidean representation for NS group (n = 30). Co-ordinates represent the relative positions created by the NS group for the 1K *Treasure Island* verbs.

Figure 6.4 shows that MDS also discerned three distinct, closely related groups of verbs in the NNS sorting task results. They can be labelled with the same names as in the case with the NS results. However, the NNS components were slightly different from those of their NS counterparts. They were:

(1) the verbs of motion of GO in the (-X, -Y) dimension (16 words): *add, remain, rest, start, sit, run, return, rise, leave, increase, fall, begin, come, go, stop* and *walk*
(2) the verbs of motion of HAVE and FIGHT in the (-X, +Y) dimension (14 words): *help, struggle, beat, fight, break, strike, die, have, make, get, keep, take, hold* and *bring*
(3) the verbs of SPEAK and THINK in the (+X, -Y) dimension (18 words): *mind, believe, feel, look, cry, listen, hear, think, stare, reply, see, ask, show, say, speak, tell, describe* and *sing.*

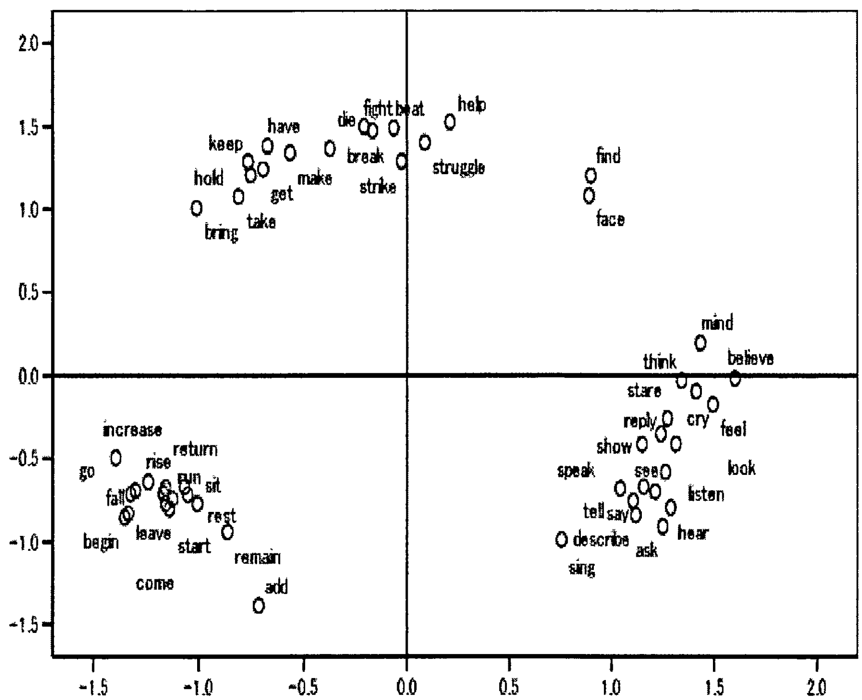

Figure 6.4. Two-dimensional Euclidean representation for NNS group (n = 30). Co-ordinates represent the relative positions created by the NNS group for the 1K *Treasure Island* verbs.

Furthermore, *find* and *face* in the (+X, +Y) dimension can be regarded as a small cluster since they are closely placed together. This cluster did not exist in the NS configuration.

Given these L1 and L2 results made by MDS as a whole, it was confirmed that MDS was effective in revealing the global structures of lexical organisation. Compared with the pathfinder analysis, MDS was more revealing in that the analysis succeeded in drawing the overall cluster structures that each participant group made in regard to the present sorting task.

6.3.3.3 Cluster analysis

Cluster Analysis was run on the matrices of NS and NNS group data. Figure 6.5 is the dendrogram of the results computed for the NS group and Figure 6.6 is the dendrogram of the results for the NNS group.

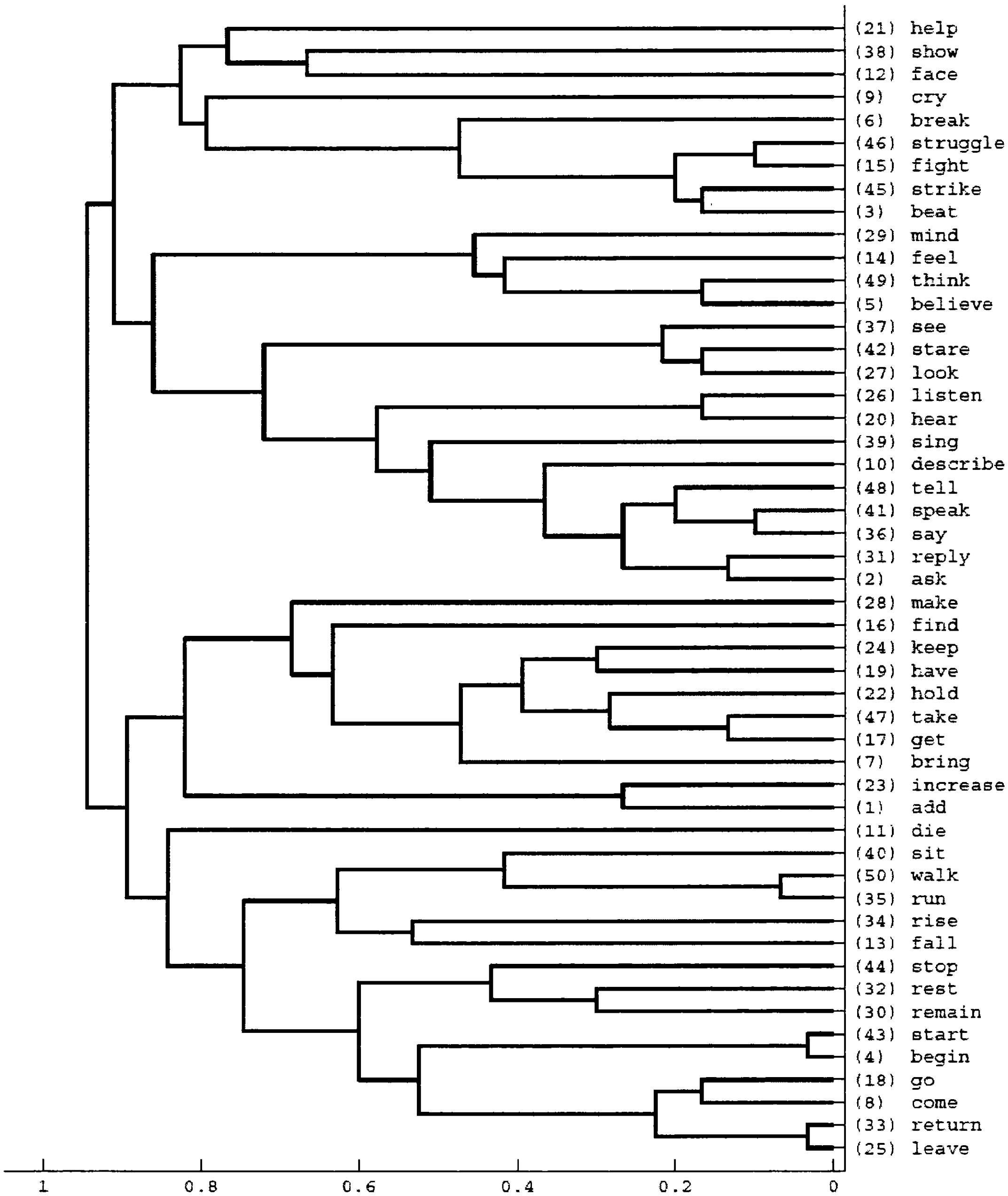

Figure 6.5. Dendrogram of NSs' sorting task results. Analysis was done by means of cluster analysis. Numbers in parentheses show the numerical order of the word in the co-occurrence matrix of Appendix 6.2a.

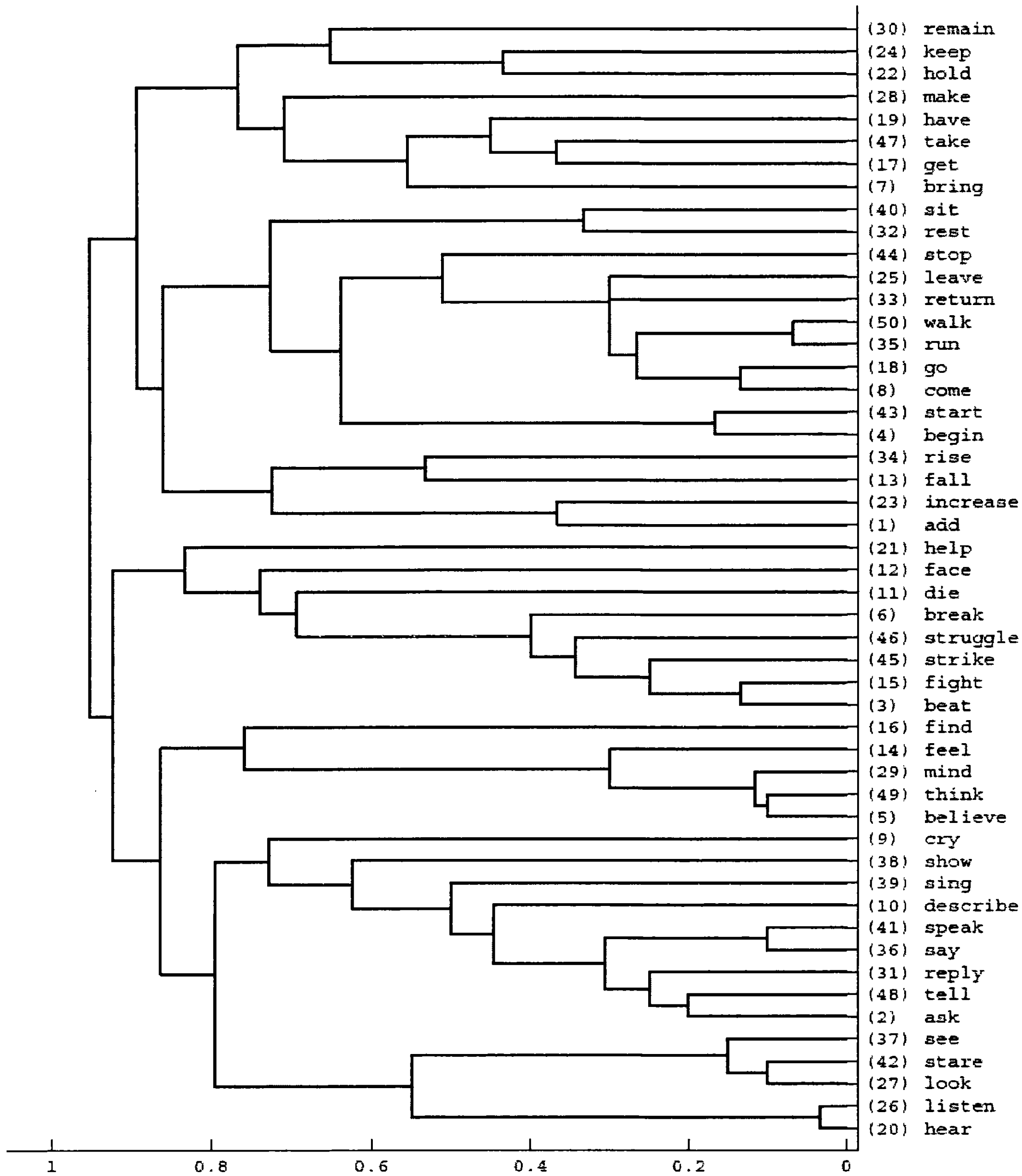

Figure 6.6. Dendrogram of NNSs' sorting task results. Analysis was done by means of cluster analysis. Numbers in parentheses show the numerical order of the word in the co-occurrence matrix of Appendix 6.2b.

Based on the dendrograms shown in Figures 6.5 and 6.6, the number of final clusters was finalised. The identification of the number of final clusters of a dendrogram was done by following three stopping rules. (a) Final clusters should range from two to 10 in number, with two to seven being the most desirable; (b) In the process of deciding a cut-off point of the dendrograms to decide final clusters, the similarity values were plotted in descending order (see Figures 6.7 and 6.8), the change in heterogeneity between clusters was assessed, and "a gap" in the plot was found; (c) Then the prior cluster solution to "a gap" was selected. The rationale is that when large increases in heterogeneity are identified in cluster analysis results, one should select the prior cluster solution because the combination joins quite different clusters. (See Hair et al. (2005), "Chapter 8, Cluster Analysis" (pp. 555-628), for the steps that should be taken to determine final cluster solutions.) The circle in Figure 6.7 shows the "gap" in the L1 group results and the circle in Figure 6.8 shows the "gap" in the L2 group results after the stopping rules above were applied to each set of data.

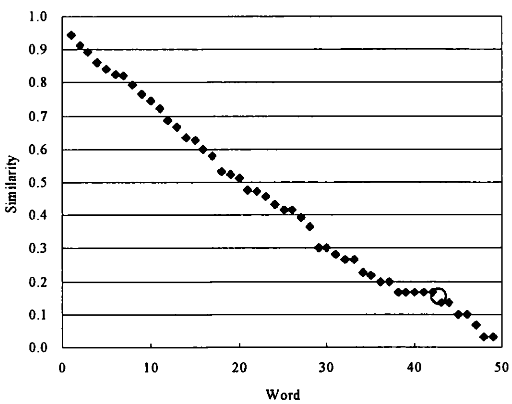

Figure 6.7. Similarity values in descending order: NS group

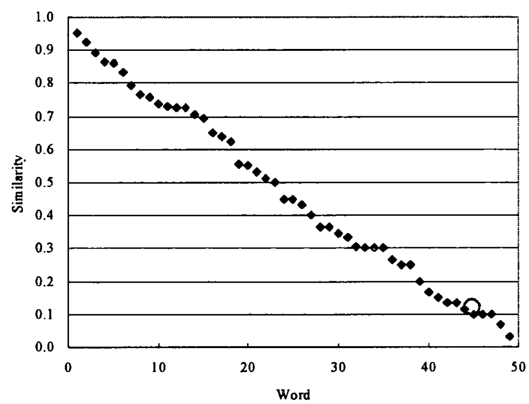

Figure 6.8. Similrity values in descending order: NNS group

Accordingly, a set of eight final clusters[1] were identified in the L1 dendrogram, whereas a set of six final clusters were discerned in the L2 dendrogram. Final clusters and lexical items composing of the clusters in the L1 and L2 dendrograms are tabulated in Tables 6.16 and 6.17, respectively. NS' final clusters were composed of the verbs of the THINK cluster (four words), the verbs of the PERCEPTION and COMMUNICATION cluster (12 words), the verbs of motion of the HAVE cluster (eight words), the verbs of motion of the INCREASE cluster (two words), the verbs of motion of the GO cluster (14 words), the verbs of the DIE cluster (one word), the verbs of the SHOW cluster (three words) and the verbs of the FIGHT cluster (six words). Meanwhile, NNS' final clusters were made up of the verbs of the THINK cluster (five words), the verbs of the PERCEPTION and COMMUNICATION cluster (14 words), the verbs of motion of the HAVE cluster (eight words), the verbs of motion of the INCREASE cluster (four words), the verbs of motion of the GO cluster (11 words) and the verbs of the FIGHT cluster (eight words). Thus, the L1 and L2 results were different from each other in the number and composition of final clusters.

---

[1] In identifying the final clusters in the L1 dendrogram, stopping rule (C) was prioritised over the objective of selecting clusters in the range of two to seven.

Table 6.16. Final clusters identified for the NS group ($k = 50$)

No.	Cluster	Word No.	Words
1	Verbs of THINK	4	*mind, feel, think, believe*
2	Verbs of PERCEPTION and COMMUNICATION	12	*see, stare, look, listen, hear, sing, describe, tell, speak, say, reply, ask*
3	Verbs of motion of HAVE	8	*make, find, keep, have, hold, take, get, bring,*
4	Verbs of motion of INCREASE	2	*increase, add*
5	Verbs of motion of GO	14	*sit, walk, run, rise, fall, stop, rest, remain, start, begin, go, come, return, leave*
6	Verbs of DIE	1	*die*
7	Verbs of SHOW	3	*help, show, face*
8	Verbs of FIGHT	6	*cry, break, struggle, fight, strike, beat*

Table 6.17. Final clusters identified for the NNS group ($k = 50$)

No.	Cluster	Word No.	Words
1	Verbs of THINK	5	*find, feel, mind, think, believe*
2	Verbs of PERCEPTION and COMMUNICATION	14	*cry, show, sing, describe, speak, say, reply, tell, ask, see, stare, look, listen, hear*
3	Verbs of motion of HAVE	8	*remain, keep, hold, make, have, take, get, bring*
4	Verbs of motion of RISE and INCREASE	4	*rise, fall, increase, add*
5	Verbs of motion of GO	11	*sit, rest, stop, leave, return, walk, run, go, come, start, begin*
6	Verbs of FIGHT	8	*help, face, die, break, struggle, strike, fight, beat*

In conclusion, all three multivariate analyses (i.e., pathfinder analysis, MDS and cluster analysis) succeeded in revealing the arrangement of and the differences in L1 and L2 lexical organisations. This is a solid evidence for Wilks and Meara's (2002) prediction that the two lexical organisations could in fact be differently arranged even if they had the same level of lexical density. Meanwhile, results of the three analyses were different from each other in important ways. Among them, the most notable difference was the degree of distinction each analysis made of the underlying cluster structures in lexical organisation. When both the L1 and L2 results were not separated but rather merged into a single network, Pathfinder identified seven major branches in the L1 results and five branches in the L2 results. MDS showed three merged major clusters of words in both the L1 and L2 results. Meanwhile, cluster analysis distinctively detected eight final clusters in the L1 results and six final clusters in the L2 results.

6.4 Discussion

Research questions addressing L1 and L2 differences in task completion time, and cluster number, size and variability were addressed in this chapter as well as in Chapters 4 and 5. These are the research questions I am posing, in addition to organisational differences in L1 and L2 cluster structures, throughout this research project. However, the central themes in this chapter were twofold. They are (a) whether the revisions made on the sorting task boosted the sensitivity in examining L1 and L2 differences and (b) which multivariate analysis (i.e., pathfinder analysis, MDS or cluster analysis) was the most reliable in examining L1 and L2 lexical organisations and their differences. These two issues will be discussed in this section.

6.4.1 Boosted task sensitivity

The revised sorting task aimed to boost the ability of the sorting task to tap more into lexical knowledge than into conceptual knowledge. To do so, in the revised sorting task, participants were directed to carry out the task as quickly as possible, whereas in the previous directions they had been allowed to take up to 20 minutes to complete the task. As reported in section 6.3, the revised directions were effective in that all participants, except one NNS participant, completed the task in under 20 minutes. L1 participants took on average 6.82 minutes less to complete the sorting task than they did the previous task, and L2 participants on average took 8.25 minutes less to complete it than they did the previous one. The differences in the results between the previous and present sorting tasks for both groups were statistically significant, showing that the revised directions functioned as hypothesised. Furthermore, utilising this "do it as quickly as possible" direction caused no substantial difference in mean task completion time between the NS and NNS groups. Although they took significantly more time to finish the previous task than their L1 counterparts, L2 participants were able to complete the current task as almost quickly as their L1 counterparts. This suggested that, at least in the case of using 1K *Treasure Island* verbs, L2 participants could access and process their L2 lexical knowledge without needing to access their L1 lexical knowledge and conceptual knowledge very much in sorting the verbs as fast as their L1 counterparts did.

Faced with this finding, it became necessary for me to modify the conclusion made on task completion time in Chapter 5. In the previous chapter, it was concluded that it would be almost impossible for L2 participants to complete a task such as a sorting task only by accessing their L2 lexical knowledge and without accessing their L1 lexical knowledge and conceptual knowledge. Considering that the L2 participants on average still took slightly more time in task completion than their L1 counterparts in the present experiment, the conclusion made on the sorting task results in the previous experiment might still be the case. However, as there was no substantial difference in task completion time between the NS and NNS groups, when they were directed to do the sorting task as quickly as they could, L2 participants seemed to have been able to complete it without much access to their L1 lexical knowledge and conceptual knowledge. Under this direction, L2 participants made more instinctive access to the meanings of L2 words and the interconnections between them. Eventually, there was no difference in task completion time between the NS and NNS groups. Thus, it was confirmed that the simple revision of "Do it as quickly as possible" in the directions for the task boosted the sensitivity to tap into L2 lexical knowledge on the part of the NNS group.

Needless to say, it is vital to give other sorting tasks to NS and NNS groups using words selected from other parts of speech (e.g., nouns and adjectives) to test whether there is no difference in task completion time between the two groups as was the case with verbs. Källkvist (1998, 1999) reported that verbs were more difficult to learn than nouns. Miller and Fellbaum (1991) and Read (2004) showed that verbs for the most part are semantically the most complex word class. A worthwhile research question to address is whether "easier" word types have developed different L2 lexical organisation in L2 mental lexicon and would produce different sorting task results, including task completion time, from the present one. This question will be addressed further in Chapters 7 and 8.

Another revision was made in the way the words tested in the sorting task were selected. Instead of randomly selecting words from the JACET 8000, the verbs for the revised sorting task were chosen from a cohesive passage of *Treasure Island*, in which the lexical items have close semantic relatedness between them. The aim of the revision was achieved in that in after-task interviews many participants commented that they could easily sort the cards into groups where they thought words were related to each other in meaning. In addition, participants who did both the previous and present sorting tasks stated that they could find more distinctive semantic relationships between the words in the revised sorting task. Moreover, participants labelled the titles of word clusters they made in their own words that corresponded to the clusters identified above in section 6.3.3. Such comments were made as: "This group is related to verbs of physical movement," "These are verbs of communication," and "This word group is concerned with fighting". Thus, the revised sorting task seemed to have greater sensitivity to tap into the underlying lexical organisation participants have than the previous sorting task did.

It should be noted that the boosted sensitivity achieved due to revisions to the task had little impact on the mean cluster number, size and variability participant groups made. Strictly speaking, both L1 and L2 participants on average made slightly bigger "largest" clusters in the present sorting task than in the previous sorting task, but the increase in both groups was not substantial. However, as the analysis above shows, the revisions made to the present sorting task led participants to more easily find the semantic relatedness between the tested words.

Participants could more easily process which words were related to each other and sort them into groups than with the previous sorting task because of the boosted semantic relatedness between the tested words. Meanwhile, owing to the distinct emptiness of the semantic relatedness of single, isolated words to other words, participants could more easily find those words and leave them as such in their sorting behaviour. However, as the results above show, the boosted sensitivity achieved by the present word selection process had little to do with L1 and L2 differences in that L2 speakers "do not yet perceive vocabulary as belonging to smaller, constrained and strongly connected sets" (Meara & Schur, 2002, p. 179). In the end, there were no distinct L1 and L2 differences in the mean "largest" cluster and the mean number of single, isolated words as well as the mean cluster number, size and variability.

6.4.2 Comparison of multivariate analyses in lexical organisation research

In this section, the validity of pathfinder analysis, MDS and cluster analysis in examining sorting task results will be discussed, and a comparison of the three analyses will be made. Finally, it will be decided which of the three is the most appropriate for the present research purpose.

Pathfinder analysis successfully revealed the local relationships of lexical items that were tapped into by the sorting task. As analysed in section 6.3.3.1, pathfinder detected seven major partitioned branches in the NS structure and five in the NNS structure. Thus the present results support what Jonassen, Beissner, and Yacci (1993) stated: pathfinder analysis "better represents local or pairwise comparisons between concepts in a knowledge domain" (p. 74) than other techniques. Meanwhile, it was not clear what the global organisations of the structures are like. Pathfinder analysis is weak at detecting the global structure of data. This suggests that pathfinder analysis requires researchers to make subjective judgment in interpreting the results (Cooke, 1990). As is often the case with subjective judgment, interpretations and conclusions might be arbitrary, and thus misleading in understanding the underlying structures.

Moreover, as Takeuchi and Utsugi (1988) indicated, the visual representation of a pathfinder associative network would be too complicated to grasp the global relationship between them if there were a number of links in the elicited data. This actually was true with the present results. There were 50 lexical items in the present experiment, and the information visualisation by pathfinder analysis contained too many pieces of elements to reach an overall conclusion. It should be noted that in Sánchez (2004) all 19 *light* verbs tested came from a single semantic field. Sánchez adopted pathfinder in analysing and visualising the data, and thus interpretation of the data was made much easier. The weakness of pathfinder analysis is derived from its calculations, in which the analysis selects merely the strongest link between tested items and all the other links are discarded. Thus, while the information visualisation of pathfinder associative networks is appealing at a glance, the representation is not based on all the data obtained by an experiment. This may not be a problem and might actually be preferred in such cases as mapping author citation frequency and network relationships (White, 2003), but it appears to be problematic with the current sorting task results. The results lack in validity if a computation discards information as pathfinder analysis does.

Furthermore, pathfinder analysis does not reveal the underlying structures of sorting task results in a way that would reflect what the task is aiming at. Although in the sorting task participants are directed to group the tested words into clusters, pathfinder does not represent the results in a clustered way. Instead, the results are represented as an associative network. As explained above, the branches are not easily compared with each other, requiring subjective judgment on the part of researchers. This is another unsuitable feature of pathfinder analysis as far as its adoption for analysing sorting task results is concerned. Thus, pathfinder analysis is not a proper technique to examine the underlying structures of lexical organisation regarding sorting task results.

Multidimensional Scaling (MDS) was superior to pathfinder analysis in examining the sorting task results in that it revealed the overall structures of lexical organisation underlying the results. Regarding applicability, MDS has an advantage when "the objective is more oriented toward understanding overall preference or perceptions rather than detailed perspectives involving individual attributes" (Hair et al., 2006, p. 662). This is a distinctive merit of MDS, which pathfinder analysis lacks because it is more oriented toward examining local structures of the data. Considering that the present project aims at examining and revealing L1 and L2 differences in cluster structures of lexical organisation by means of sorting tasks, it is clear that MDS is more appropriate than pathfinder analysis.

However, MDS has two serious flaws that detract from its usefulness. One is concerned with the fact that MDS gives no explicit clustering information among tested attributes. This is because MDS only provides us with "the perceived relative image of a set of objects" (Hair et. al., 2006, P. 632). Rapoport and Fillenbaum (1972) and Routh (1994) reported this weakness of MDS when they examined sorting task data by MDS (see Chapter 2). This held true with the present sorting task results as well. As confirmed above, MDS showed that both the NS and NNS groups were shown to have three major clusters: the verbs of motion of GO, the verbs of motion of HAVE and FIGHT, and the verbs of SPEAK and THINK. The acute problem lies in the second cluster, the verbs of motion of HAVE and FIGHT, where the verbs of HAVE (e.g., *bring, get, hold, find*) and the verbs of FIGHT (e.g., *strike, struggle, beat, fight*) are merged together into a single tight cluster. This merged cluster is difficult to interpret easily and does not seem to be real in lexical organisation. It would have been more instinctively valid if the clusters had been identified to be separate from each other. Another problem is closely related to the first one. As in the case with pathfinder analysis, MDS requires researchers to make subjective judgment and such judgment "is not easy and sometimes misleading" (Chen, 2003, p. 155). Regarding the verbs of motion of the HAVE and FIGHT cluster, we have this difficulty in interpretation. If a study were attempting to examine the global structures of L1 and L2 lexical organisations, it would be easy to understand that the verbs of HAVE and those of FIGHT are similarly tightly clustered in both NS and NNS lexical configurations. Meanwhile, if an attempt were to examine L1 and L2 differences in cluster structures of lexical organisation, such an MDS result as the present one, where both NS and NNS had three major identical clusters and one of them was difficult to interpret, would be insufficient. Given these shortcomings, MDS is not the most appropriate selection in analysing the sorting task results.

Considering the weaknesses of pathfinder analysis and MDS identified above, cluster analysis appears to be the most reliable multivariate analysis in examining L1 and L2 differences in lexical organisation regarding the sorting task results. There are two reasons that account for the validity of cluster analysis. First, the information visualisation as a final result of cluster analysis straightforwardly represents what a sorting task aims at revealing in lexical organisation. A sorting task directs participants to sort words into clusters. Cluster analysis clearly shows the results of analysis in a way that represents the underlying cluster structures tapped into by a sorting task. Second, by analysing dendrograms by means of stopping rules to finalise the number of final clusters, it is easy to decide objectively how many distinctive word clusters were made. Regarding the present sorting task, cluster analysis revealed eight underlying cluster structures in the NS results and six in the NNS results. The two groups shared five of the final clusters (verbs of THINK, verbs of PERCEPTION and COMMUNICATION, verbs of HAVE, verbs of GO and verbs of FIGHT) with each other, but the components were different from each other (see Tables 6.16 and 6.17). Moreover, the two groups had different final clusters for the rest of the structures. Particularly, the NS group made a final cluster of verbs of DIE where *die* was the only component, whereas *die* was grouped into the verbs of the FIGHT cluster (which had eight lexical items) in the NNS results. This is what pathfinder analysis and MDS failed to detect.

It should be noted that both Rapoport and Fillenbaum (1972) and Routh (1994) first ran MDS on their sorting task results, but the final products were not clear-cut in grasping the features of the cognitive structures they addressed. MDS could not reveal the distinct clusters underlying their sorting task results. Eventually, both studies further analysed the data by applying cluster analysis to them and revealed the distinctive features in the organisation they addressed. Routh (1994), which addressed the semantic nature of quantifiers in cognitive structures, stated that by cluster analysis "most quantifiers are well defined in terms of unique features, but they also combine into clusters that are fairly interpretable" (p. 209). Miller (1969) adopted cluster analysis as well to examine sorting task results regarding nouns in memory and succeeded in uncovering the structures (see Chapter 2). This is also true with the present sorting task where cluster analysis revealed distinct differences in the L1 and L2 lexical organisations of 1K verbs.

Given the comparison of the three multivariate analyses above as a whole, cluster analysis is the most appropriate choice in investigating L1 and L2 structural differences in lexical organisation. Thus, the comparison made in this chapter confirmed the conclusion that the literature review in Chapter 2 made, which was that cluster analysis is the most appropriate method to analyse sorting task results. In cluster analysis, data is calculated in a way that puts individuals or groups into clusters in view of the distance between the tested objects. Considering that the present research project aims at examining L1 and L2 differences in cluster structures in the mental lexicons using sorting tasks, cluster analysis is the most reliable calculation to examine and represent the underlying structures.

Cluster analysis is a group-oriented analysis, as is the case with other types of multivariate analyses. Thus, it would be extremely difficult to investigate the differences of individual participants in the results by means of cluster analysis. This is a concern in particular when analysing whose results have more variability (or are more consistent), the NS or NNS group. In short, is it possible to examine cluster analysis results in terms of whether L1 participants are less varied than their L2 counterparts as Meara and Schur and other word association test-based studies revealed? This is a question to be posed in future experiments. Because of this limitation of cluster analysis, it would be desirable to run descriptive statistics-oriented analysis (e.g., *t*-tests, *F*-tests and ANOVAs) which would address individual differences in cluster number, size and variability in the sorting task results of future experiments as well.

6.5 Conclusion

In the present revised sorting task, participants were directed to complete the task as quickly as they could. The verbs tested were selected from a cohesive passage in *Treasure Island.* These revisions boosted task sensitivity in that both the NS and NNS groups took significantly less time to complete the task than they did with the previous sorting task. In addition, both groups were not significantly different from each other in mean completion time. This shows that L2 participants activated their L2 lexical knowledge more than their L1 lexical knowledge and conceptual knowledge in completing the verb sorting task.

An interesting question then arises: Will other word classes (e.g., nouns, adjectives, etc.)

produce similar results in a sorting task experiment to those verbs have produced? If that is the case, it means that L2 participants can access and process lexical knowledge instinctively as quickly as their L1 counterparts do. If it is not, the network structures of lexical organisation that Aitchison (1994, 2003) and Meara (1997, 2002, 2004) postulate are not similarly arranged and are different from each other depending on the word type. These are the issues to be addressed in Chapters 7 and 8.

Similarly, this research project has not yet reached a conclusion on L1 and L2 differences in cluster number, size and variability. Similar to the sorting task results reported in previous chapters, in the present experiment, the NS group tended to make a larger number of clusters, smaller cluster size and less variability than NNS group as in Meara & Schur (2002). However, in most cases the differences were not statistically significant. So far, randomly selected 0.5K words, randomly selected 1K verbs and 1K *Treasure Island* verbs were tested, and only the sorting task using 1K *Treasure Island* verbs had boosted sensitivity in examining lexical organisation. Other sorting tasks that have boosted task sensitivity using other word types must be tested to understand the overall picture of L1 and L2 lexical organisations in mental lexicons. Particularly, will the finding that Meara and Schur made be supported by the sorting task results and generalised? Or does a sorting task tap into a different realm of lexical organisation from the word association test-based study? These questions will be addressed in the following chapters.

Multivariate analyses (i.e., pathfinder analysis, MDS and cluster analysis) showed that there were actually structural differences between L1 and L2 lexical organisations. Through these analyses, cluster structures of L1 and L2 lexical organisations were revealed and the differences were confirmed. This was not achieved by searching for small-scale "native-like links" in the sorting task results reported in Chapter 5. Comparison of the analyses showed that cluster analysis appears to be the most reliable multivariate analysis method to tap into overall L1 and L2 lexical organisations and the differences produced by sorting tasks. In Chapter 7, a non-verb word class will be tested by means of the present revised sorting task and the results will be analysed by means of cluster analysis while also keeping in mind the present research questions as well.

By solely examining the L1 and L2 dendrograms that the cluster analysis made, it was extremely difficult to investigate whether the two groups are different from each other in cluster variability. However, to validate cluster analysis in examining the sorting task results, further investigation is needed. That is, the question of whether L1 lexical organisation is more consistent than its L2 counterpart regarding sorting task results should be posed. This issue will be addressed in Chapter 7 as well.

Chapter 7: Cluster variability in L1 and L2 lexical organisations

7.1 Introduction

In chapter 6, we directed participants to complete the sorting task as quickly as possible. We found that the revised task functioned effectively and L2 participants were able to complete the task as fast as their L1 counterparts. Due to this direction in the task, L2 participants activated and processed their L2 lexical knowledge more than their conceptual knowledge and L1 lexical knowledge in completing the task. However, as in the previous experiment, there was no perceived significant difference between the NS and NNS groups in cluster number, size and variability. Thus the present project has not yet achieved conclusive results on the structural differences in L1 and L2 lexical organisations. This might be accounted for by the prediction Wilks (1999) made that the organisational differences between L1 and (advanced) L2 speakers are more subtle than one would expect. Deese (1964) argues that "it is appropriate to study the structure of associative meaning within a grammatical class [word class] and to relate the results to the organization of that class" (p. 347). Thus, another possible explanation is that word class might be an important factor in L2 vocabulary acquisition, but that the differences in verbs are the least noticeable. Which prediction holds true with what L1 and L2 lexical organisations are actually like? The question has not yet been answered. It is worthwhile to test this question in the rest of this project, and the sorting task results using 1K adjectives will be reported in this chapter.

By the introduction of multivariate analyses (i.e., pathfinder analysis, MDS and cluster analysis) in the data analysis of the previous chapter, it was revealed that L1 lexical organisation was qualitatively different from its L2 counterpart. Furthermore, by comparing the three analyses, we found that cluster analysis is the most reliable in addressing how lexical items are arranged in mental lexicons regarding the data collected by a sorting task. Cluster analysis showed that NS made eight final clusters of words while NNS made six. Furthermore, their components were different from each other. The cluster structures identified in the dendrograms of the two participant groups were found to be organised in a clearly different way. A pertinent problem, however, lies in the fact that these findings are concerned with group data analysis, not with individual participant data analysis. This is particularly related to the issue of whether L1 participants are more consistent in their lexical organisation than L2 participants are. In other words, in regard to cluster analysis, are the dendrograms of NS results less varied than NNS results as Meara and Schur (2002) and other previous word association test-oriented studies (Fitzpatrick, 2006; Meara, 1978, 1983; Postman & Keppel, 1970; Riegel & Zivian, 1972; Szalay & Deese, 1978) reported? Are the two lexical organisations more similar to than different from each other as shown by the subtle differences in cluster number, size and variability detected by the sorting tasks in the previous chapters? Furthermore, regarding group data analysis, are the L1 and L2 group dendrograms substantially different from each other? In order to answer these questions we need to further examine the cluster structures of L1 and L2 lexical organisations.

As confirmed in the previous chapter, cluster analysis has the power to answer the question of L1 and L2 differences in lexical organisation. But even with cluster analysis we still cannot answer the question of whether L1 lexical organisation is less varied than its L2 counterpart. Cluster analysis is primarily concerned with group data as is the case with other multivariate analyses. The results provide us with no information regarding which group (the NS or NNS group) is more consistent in sorting behaviour among the participants. Thus, it is necessary to introduce another statistical technique that can answer the question of cluster variability in L1 and L2 lexical organisations.

In the previous experiment using 1K *Treasure Island* verbs, we confirmed that L2 participants accessed and processed their L2 lexical knowledge as fast as their L1 counterparts in completing a sorting task. In this chapter, this finding will be tested using 1K adjectives (research question 1). Meara and Schur (2002) argued that L2 speakers, for the most part, "do not yet perceive vocabulary as belonging to smaller, constrained and strongly connected sets and they tend to associate words in more diverse and less predictable ways than native speakers do" (p. 179). This is based on their finding that L1 participants in WAT results had more clusters, smaller cluster size and less variability than L2 participants. As stated above, the present project has failed thus far to support their finding. Hypothesising that there might be structural differences in cluster structures between different word classes, this issue (research question 2) will be addressed in this chapter again. The third research question is related to cluster variability of L1 and L2 lexical organisations. We will attempt to answer this question by running a permutation test to examine the dispersion of goodness-of-fit of individual participants versus the group data. This will be explained further in 7.2.3 Data Analysis.

Meara (2007a), through a computer simulation study of a small lexicon, revealed that variations in local structure are less significant than might be expected, and suggested that the overall number of connections between words in the lexicon is the principal factor that affects the outcome of simulations. He also indicated that local structure on a small scale is relatively unimportant. Meara's research is vital in that it leads us to attend to the importance of the global structure of L1 and L2 mental lexicons and their organisational differences. Wilks and Meara (2002) predicted that two [L1 and L2] networks with the same density "could in fact be quite differently arranged in terms of how the connections between points [lexical items] are disposed." (p. 319). In the sorting task using 1K verbs, firm evidence for their prediction was obtained. Will we find further evidence for it with a sorting task using 1K adjectives? This issue will also be addressed in this chapter (research question 4).

The above-mentioned points are summarised below:

1. Do L2 participants complete the sorting task as quickly as L1 participants?
2. Do L1 participants make a larger number of clusters and fewer words per cluster than their L2 counterparts do while the results L1 participants produce are less varied than their L2 counterparts?
3. Is the L2 lexical organisation less consistent than the L1's, where the sorting task results of L2 participants are more varied than their L1 counterparts?
4. Is the L2 lexical organisation structurally different from its L1 counterpart?

7.2 Method

7.2.1 Participants

This study had two participant groups. The first was comprised of 30 adult, native speakers of English (NS). They were teachers of English in Kumamoto, visiting scholars and students studying at Kumamoto University. The second consisted of 30 adult, advanced-level Japanese speakers of English (NNS). They were either English teachers at the college level or persons having a high competence of English as judged by a TOEFL score of 213 or more on the computer-based version or a score of 550 or more on the paper-based version or a TOEIC score of 730 or more that had been taken within the last two years.

7.2.2 Data Collection

Fifty adjectives were used in the present sorting task. Using the procedure followed in the verb sorting experiment, 50 adjectives were randomly selected from among all the adjectives contained in Robert Stevenson's (1883), *Treasure Island*, "Chapter 32. The Treasure-hunt — The Voice Among the Trees". The 50 adjectives (which were among the first, 1000 high frequency words from the JACET 8000) were as follows: *afraid, alone, back, bad, big, black, blue, certain, clear, close, dead, deep, far, fine, forward, good, great, half, happy, hard, high, hot, kind, large, last, light, little, long, low, natural, near, nice, old, open, ready, recent, red, right, round, single, small, strange, strong, successful, sure, true, well, whole, wide, wrong.*

After piloting, the sorting task (which was of the same design as the sorting task reported in Chapter 6 except for the type of words tested) was administered to each of the participants individually.

7.2.3 Data Analysis

To answer research questions 1 and 2, the sorting task data was analysed by comparing the means and *SD*s of L1 and L2 results as were done in the previous experiments. This was performed by means of a *t*-test and an *F*-test on the time taken to complete the task, cluster number and size, the number of single, isolated words, and the "largest" cluster participants made.

To address research question 3 of whether the sorting task results of participants of the NNS group were more varied than those of the NS group, a permutation test of individual differences between groups was conducted. Permutation tests make use of distance matrices, where a distance matrix is the sum of the lengths (i.e., heights) of each individual participant's dendrogram trees. Then each individual participant's distance matrix data was examined to find its degree of goodness-of-fit with the group (L1 or L2) dendrogram. In the examination, the smaller a value is, the better an individual result fits with the group result (i.e., the dispersion between them is smaller). The degree of goodness-of-fit of an individual distance matrix with the group data (distance matrix) is found by calculating the square of the Minkowski distance (a measure between two distance matrices which is the most valid for sorting task results). The Minkowski distance is computed as follows:

$$d(T_1,T_2)=\left\{\sum_{i,j}\left|\tilde{D}_1(i,j)-\tilde{D}_2(i,j)\right|^2\right\}^{\frac{1}{2}}$$

where d (T_1, T_2) is the distance between T_1 and T_2;

T_1 and T_2 are dendrograms;

$\tilde{D}_1(i,j)$ is the distance between x_1 and x_2 on T_1;

$\tilde{D}_2(i,j)$ is the distance between x_1 and x_2 on T_2.

The average of L1 results was compared to that of L2 results on the assumption that the results of individual participants of the two groups would not be different from each other.
The sorting task results of all participants (i.e., 30 NS and 30 NNS participants in the present experiment) were mixed in a random way and divided (permutated) into two new groups. Eventually, they formed two dummy groups of data consisting of 30 participants each. For each group of dummy data, a dummy dendrogram was produced. Then the Minkowski distance between the distance matrices of the pair of dummy dendrograms was calculated. This procedure was repeated a sufficiently large number of times[1] (e.g., 1,000 times). A

[1] For the present experiment, a permutation test was administered 5,000 times. A reliability test

histogram that represented the data distribution was produced, using these Minkowski distance data sets, and it was then compared to the average Minkowski distance between the original (real) distance matrices of individual participants' dendrograms of the NS and NNS groups. If the average Minkowski distance turns out to be an outlier in the data distribution of the histogram, then the individual participants' difference between the NS and NNS groups is declared to be statistically significant. This is the way a permutation test works. It should be noted that a permutation test has an advantage over a *t*-test and an *F*-test in that the test has the power to compare such values as the ones that are obtained as a result of complicated calculations of cluster analysis. (See Saito & Yadohisa, 2006; Takemura, 1991, for details of cluster analysis and permutation tests).

To address research question 4, two types of statistical tests were given. First, while addressing the issue of individual differences of the L1 and L2 sorting task results, a *t*-test on individual participants' distance matrices of both groups was run, where the distance matrix was a sum of lengths of the participant's dendrogram trees. Second, a permutation test was run on the NS and NNS group data of distance matrices to answer the question of whether the two group dendrograms were significantly different from each other. In addition, L1 and L2 group dendrograms as a final graphic representation of cluster analysis were depicted and their clustering components (lexical items) were compared to each other to explore the qualitative differences in their cluster structures.

7.3 Results

7.3.1 Time taken to complete task

Table 7.1 shows the means and *SD*s of the amount of time the NS and NNS groups took to complete the sorting task using the 50 high frequency adjectives selected from *Treasure Island.*

Table 7.1. Time to complete 1K *Treasure Island* adjectives sorting task

	NS ($n = 30$)	NNS ($n = 30$)	*t*-value	*F*-value
Mean	9.96	10.64	0.80n.s.	1.00n.s.
SD	3.31	3.32		

Note. n.s. = not significant.

On average, NNS took 0.68 minutes (41 seconds) more to complete the task than NS did. An unmatched *t*-test revealed that there was no substantial difference between the two means

addressing whether 5,000 times was large enough for the data size of the present experiment was run. The result showed that 5,000 times was large enough to get reliable results. Calculation of a reliability test and that of permutation tests were conducted by means of MATLAB (Matrix Laboratory) Version 7.5.0.342 (R2007b) (The MathWorks, Inc., 2007).

($t(58) = 0.80$, n.s.). There was no participant in either group who took more than 20 minutes to complete the task (the maximum among the NS was 16.4 minutes and the maximum among the NNS was 17.1 minutes). Given both the results of the previous and present sorting tasks regarding task completion time, it was suggested that L2 participants completed the sorting tasks almost as quickly as their L1 counterparts.

7.3.2 Cluster number, size and variability

This section will answer research question 2 addressing L1 and L2 differences in cluster number, size and variability. Analysis will be made by comparing the NS and NNS groups regarding the mean number of clusters, mean number of words per cluster, mean number of single, isolated words, and the mean largest cluster participants made.

Table 7.2 shows the mean number of clusters participants made in the present sorting task, where the count excludes single, isolated words. The table reveals that on average NS produced a slightly larger mean cluster number than NNS did, but the difference was not statistically significant (t (58) = 0.73, n.s.). As the *SD*s show, the variance of the NS was slightly larger than that of the NNS. However, an F-test showed that there was no meaningful difference in variability between the two groups ($F(29) = 0.74$, n.s.). Thus, the present sorting task failed to distinguish NS from NNS in mean cluster number and the variability generated.

Table 7.2. Mean number of clusters (which excludes single, isolated words)

	NS ($n = 30$)	NNS ($n = 30$)	t-value	F-value
Mean	8.73	8.27	0.73n.s.	0.74n.s.
SD	2.65	2.27		

Note. n.s. = not significant.

Table 7.3 shows the mean number of words per cluster participants in the present sorting task made, where the count excludes single, isolated words. The table reveals that, on average, NS produced a slightly smaller number of words per cluster than NNS did, but the difference was not statistically significant (t (58) = 0.60, n.s.). As the *SD*s show, the variances of the two groups were almost identical and an F-test revealed that there was no substantial difference in variability between the two groups ($F(29) = 1.04$, n.s.). The present sorting task failed to distinguish NS from NNS in regard to the mean number of words per cluster and the variability that participants produced.

Table 7.3. Mean number of words per cluster (which excludes single, isolated words)

	NS ($n = 30$)	NNS ($n = 30$)	t-value	F-value
Mean	5.81	6.10	0.60n.s.	1.04n.s.
SD	1.92	1.97		

Note. n.s. = not significant.

Table 7.4 shows the mean number of single, isolated words that participants in the present sorting task produced. The table reveals that on average NS generated a slightly larger mean number of single, isolated words than NNS did, but the difference was too slight to reach a statistically significant level ($t(58) = 0.38$, n.s.). Thus, the number of single, isolated words did not have a profound impact on the overall results and failed to produce any differences between the two groups.

Table 7.4. Mean number of single, isolated words

	NS ($n = 30$)	NNS ($n = 30$)	t-value	F-value
Mean	3.80	3.45	0.38n.s.	1.40n.s.
SD	3.36	3.42		

Note. n.s. = not significant.

Table 7.5 shows that on average L1 participants generated a smaller "largest" cluster than their L2 counterparts. However, an unmatched t-test revealed that there was no statistically significant difference between the two means ($t(58) = 1.09$, n.s.). Thus, the production of a "largest" cluster did not bring about a noticeable effect in regard to L1 and L2 differences in sorting behaviour.

Table 7.5. Mean largest cluster participants made

	NS ($n = 30$)	NNS ($n = 30$)	t-value	F-value
Mean	11.67	12.77	1.09n.s.	0.630n.s.
SD	4.33	3.44		

Note. n.s. = not significant.

Given these results as a whole, the present sorting task failed to produce distinct L1 and L2 differences regarding cluster number, size and variability, which is the same result as the previous sorting task.

7.3.3 Comparison of L1 and L2 dendrograms: Individual dispersion of goodness-of-fit against group data

This section will answer research question 3 of whether the L2 lexical organisation is less consistent than L1's and if the sorting task results of L2 participants is more varied than their L1 counterparts. For the investigation, I computed the degree of goodness-of-fit of individual NS participants' dendrograms against the group's dendrogram as calculated by the square of the Minkowski distance between the distance matrices (= dT_NS), and the degree of goodness-of-fit of individual NNS (Japanese) participants' dendrograms against the group's dendrogram as calculated by the same test (= dT_JP). See Appendix 7.1 for the tabulated results of each calculation. Table 7.6 shows the T values (= average dispersion of dendrograms of individual participants against the dendrogram of the NS and NNS results, respectively). In the table, the p-value shows whether the difference between the results of the NS and NNS is statistically significant. The T value of NS fitted better with the group result than that of NNS. However, the difference did not reach a statistically significant level ($p = 0.413$). This shows that the NNS individual results were not substantially more varied than the NS individual results.

Table 7.6. Permutation test result: Average dispersion of dendrograms of individual participants against the dendrogram of the group data

	NS ($n = 30$)	NNS ($n = 30$)	p-value
T	103.66	106.21	0.413

Note. T = Average dispersion of dendrograms of individual participants against the dendrogram of the group data; p-value = significant level of the difference of the T value between the NS and NNS results as calculated by a permutation test which was administered on the assumption that the average dispersion of the NS group's individual dendrograms and the NNS's would be equal to each other.

7.3.4 Organisational differences of L1 and L2 cluster structures

This section will answer research question 4 addressing whether the L2 lexical organisation is structurally different from its L1 counterpart, which is something that should be revealed in the sorting task results. Analysis was done in two ways. First, while addressing the issue of individual differences of L1 and L2 sorting task results (where single, isolated words were excluded), a t-test was run on the individual participants' distance matrices for both groups, where a distance matrix is the sum of the length of dendrogram trees (see Appendix 7.2). Table 7.7 shows the means and *SD*s of the distance of each individual participant's dendrogram (distance matrix) for both the NS and NNS groups.

Table 7.7. Distance of each individual participant's dendrogram

	NS (n = 30)	NNS (n = 30)	t-value	F-value
Mean	7.73	7.27	0.73n.s.	0.74n.s.
SD	2.65	2.27		

Note. n.s. = not significant.

Table 7.7 shows that the result of the NS group was bigger than their NNS counterpart's, suggesting that on average L1 participants made clusters having a longer distance than their L2 counterparts. However, the result was not statistically significant (t (58) = 0.73, n.s.). Thus, the cluster analysis performed on the distances of individual participant's dendrograms did not reveal a marked difference between the two groups.

Second, cluster analyses were run on the matrices of NS and NNS group data (Appendices 7.3a and 7.3b). Figure 7.1 is the dendrogram of the results computed for the NS group and Figure 7.2 is the dendrogram of the results for the NNS group.

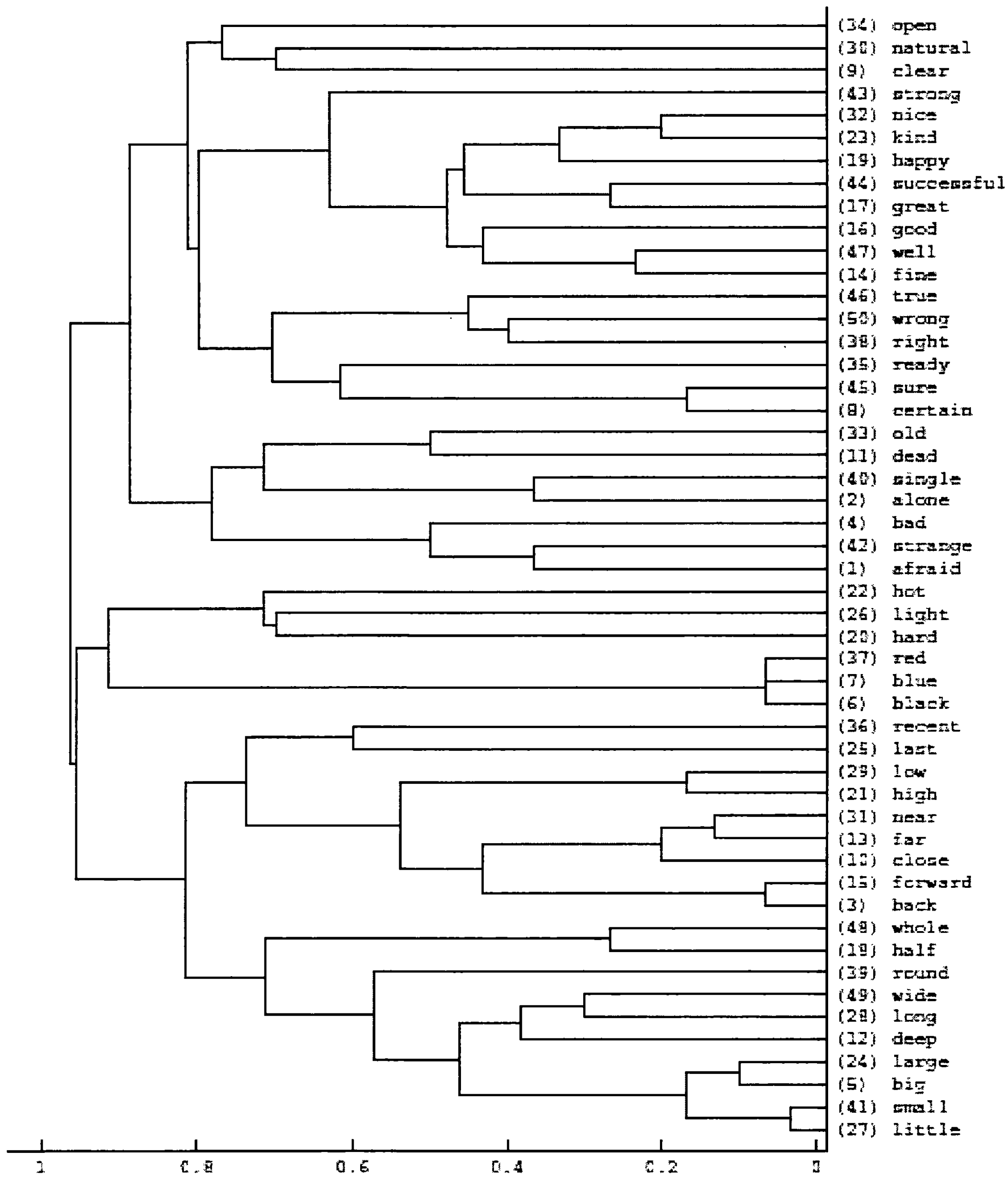

Figure 7.1. Dendrogram of NS's sorting task results. Analysis was done by means of cluster analysis. Numbers in parentheses show the numerical order of the word in the co-occurrence matrix of Appendix 7.3a.

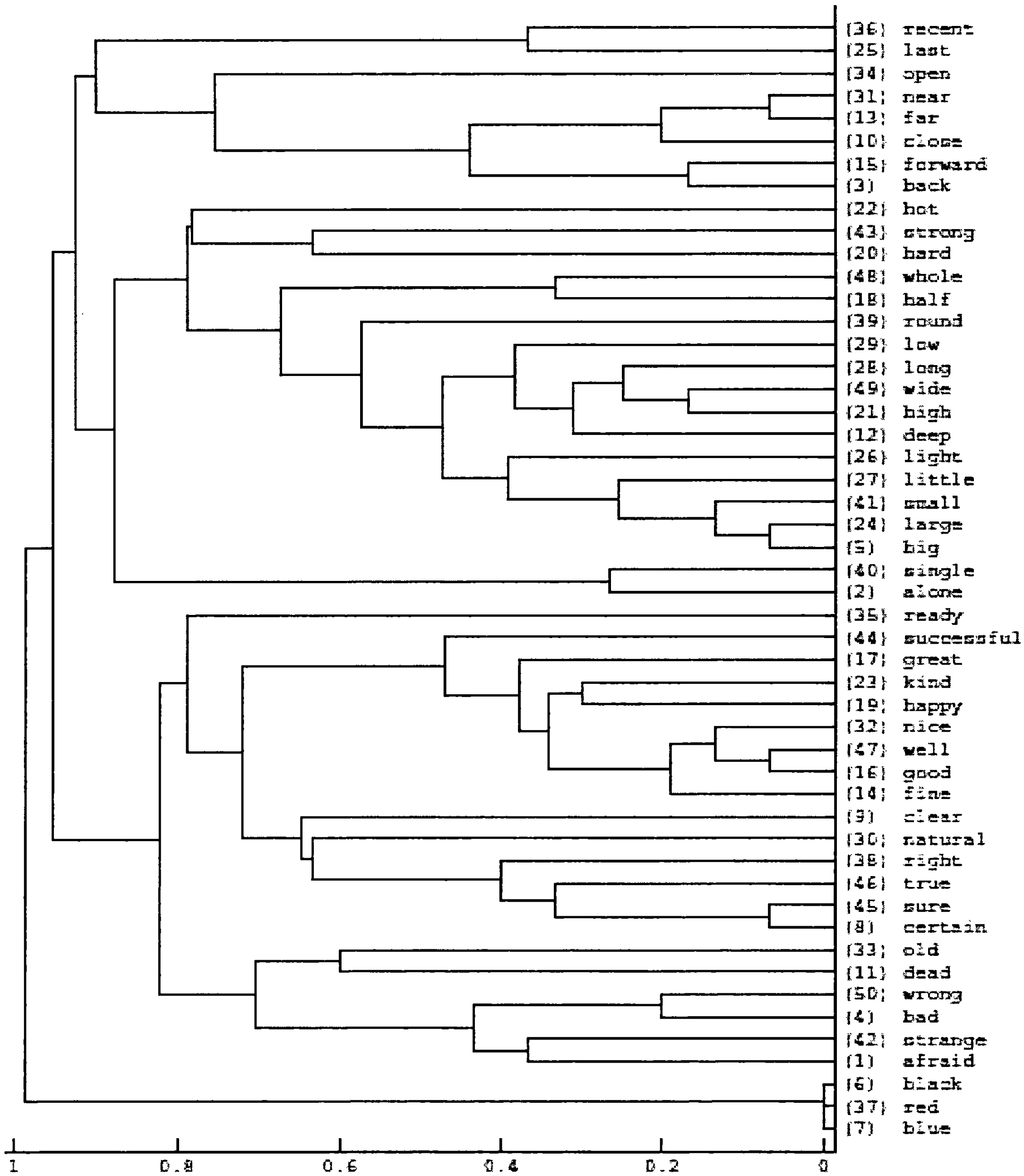

Figure 7.2. Dendrogram of NNS's sorting task results. Analysis was done by means of cluster analysis. Numbers in parentheses show the numerical order of the word in the co-occurrence matrix of Appendix 7.3b.

Based on the dendrograms shown in Figures 7.1 and 7.2, the number of final clusters was finalised. As was done in Chapter 6, the identification of the number of final clusters of a dendrogram was carried out by following stopping rules. As a result, a set of five final clusters were located in the L1 dendrogram, whereas a set of three final clusters were identified in the

L2 dendrogram. It should be noted that the present sorting task results using 1K adjectives are consistent with previous results using 1K verbs in that the NS group made a larger number of clusters than the NNS group did in both experiments. Final clusters and lexical items composed of the clusters in the L1 and L2 dendrograms are shown in Tables 7.8 and 7.9, respectively. NS' final clusters were composed of the "Positive meanings and emotions" cluster, the "Negative meanings and emotions" cluster, the "Physical dimension of degrees" cluster, the "Physical dimension of time, space, shape and size" cluster, and the "Physical dimension of colours" cluster. Meanwhile, NNS' final clusters were comprised of the "Positive and negative meanings and emotions" cluster, the "Physical dimension of time, space, degrees, shape and size" cluster, and the "Physical dimension of colours" cluster. Thus, the NS and NNS clustering results turned out to be distinctively different from each other in the number and composition of final clusters.

Table 7.8. Final clusters identified for the NS group ($k = 50$)

No.	Cluster	Word No.	Words
1	Positive meanings and emotions	18	*open, natural, clear, strong, nice, kind, happy, successful, great, good, well, fine, true, wrong, right, ready, sure, certain*
2	Negative meanings and emotions	7	*old, dead, single, alone, bad, strange, afraid*
3	Physical dimension of degrees	3	*hot, light, hard*
4	Physical dimension of time, space, shape and size	19	*recent, last, low, high, near, far, close, forward, back, whole, half, round, wide, long, deep, large, big, small, little*
5	Physical dimension of colours	3	*black, blue, red*

Table 7.9. Final clusters identified for the NNS group ($k = 50$)

No.	Cluster	Word No.	Words
1	Positive and negative meanings and emotions	21	*ready, successful, great, kind, happy, nice, well, good, fine, clear, natural, right, true, sure, certain, old, dead, wrong, bad, strange, afraid*
2	Physical dimension of time, space, degrees, shape and size	26	*recent, last, open, near, far, close, forward, back, hot, strong, hard, whole, half, round, low, long, wide, high, deep, light, little, small, large, big, single, alone*
3	Physical dimension of colours	3	*black, blue, red*

A permutation test was run on the L1 and L2 group data of distance matrices to test whether the L1 dendrogram represented in Figure 7.1 would be different from the L2 dendrogram represented in Figure 7.2. The test confirmed that the two dendrograms differed significantly from each other at the 1% significance level ($p = 0.0012$). This further confirms that L1 lexical organisation is markedly different from L2 lexical organisation.

7.4 Discussion

In the present sorting task using 1K adjectives under the direction of "Do it as quickly as possible", there was no statistically significant difference in task completion time between the NS and NNS groups. It was confirmed that L2 participants could activate and process their L2 lexical knowledge as fast as their L1 counterparts under the task direction as was the case with the previous experiment using 1K verbs. Meanwhile, given the results of the previous and present sorting tasks, there appear to be within-group differences depending on the word types used in the sorting tasks. This issue will be addressed first in this section. A permutation test failed to reveal that the NNS results were statistically significantly more varied than the NS results while the NS results fitted better with the group results than the NNS results did. The experiment also failed to detect any statistically significant difference between the mean distances of individual participant's dendrograms for the NS and NNS groups. However, when the test was run on the group data of distance matrices, it was discovered that L2 lexical organisation was distinctly different from their L1 counterpart. At first glance, these results seem to contradict each other. Thus, as a second discussion point, the relationship between the individual dendrograms and the group dendrograms will be addressed. Lastly, evidence for organisational differences between L1 and L2 mental lexicons will be discussed.

7.4.1 Time taken to complete task: Comparing the verb and adjective sorting task results

This section addresses the question of whether both the NS and NNS groups carried out the sorting tasks at a similar speed, irrespective of the types of the words tested. In other words, did L1 and L2 participants similarly activate and process their English lexical knowledge, no matter which type of words were used in the experiments?

To answer this question, I compared the results of the present experiment with those of the previous one that was administered under the same directions but used a different type of words. In both experiments, no statistically significant difference was detected between the NS and NNS groups in regard to the time taken to complete the task. However, as suggested above, both participant groups seem to have taken more time to complete the sorting task using adjectives (the present experiment) than they did with verbs (the previous experiment). To examine whether participants of each group actually took more time in completing the sorting task using adjectives than using verbs, unmatched t-tests were run on the data. Table 7.10

shows a comparison between the verb and adjective experiments regarding the NS results, and Table 7.11 shows the comparison in regard to the NNS results.

Table 7.10. NS time to complete the sorting task: Comparison between verb and adjective experiments

	Verbs ($n = 30$)	Adjectives ($n = 30$)	t-value	F-value
Mean	7.94	9.96	2.37*	1.02n.s.
SD	3.28	3.31		

Note. * $p < .05$; n.s. = not significant.

Table 7.11. NNS time to complete the sorting task: Comparison between verb and adjective experiments

	Verbs ($n = 30$)	Adjectives ($n = 30$)	t-value	F-value
Mean	9.95	10.64	1.17n.s.	0.73n.s.
SD	3.89	3.32		

Note. n.s. = not significant.

Table 7.10 shows that the NS group took 2.02 minutes more when they did the adjective sorting task than when they did the verb sorting task and that the result was statistically significant ($t(58) = 2.37, p < 0.05$). Meanwhile, Table 7.11 shows that for the NNS group, the difference (= 0.69 min.) was not significant.

Thus, it was confirmed that L1 participants needed a different length of time to complete sorting tasks which used words taken from different word classes. For native speakers of English, adjectives took more time to sort into clusters according to meanings than verbs did. This was not the case with L2 participants. They had a slight tendency to take more time with adjectives than with verbs as did L1 participants, but the difference was not statistically significant. Given these results, it is hypothesised that L1 English speakers are more sensitive than L2 English speakers in working out the meaning of relationships between words depending on the word type they are sorting. However, this is only comparing the results of two word classes (i.e., verbs and adjectives). To confirm this hypothesis, I need to carry out a further experiment with another word set taken from a third word class under the same task directions as the first and second ones. This issue will be addressed in Chapter 8.

7.4.2 Relationship between individual dendrograms and group dendrograms

The results reported in section 7.3 showed that there was no distinct difference between the L1 and L2 dendrograms at the individual participants' level, but there was a statistically significant difference between the L1 and L2 group dendrograms. This section addresses this somewhat unexpected result concerning the relationships between the individual and group results. We will discuss three points.

First, 1K adjectives, which were selected from a cohesive passage of *Treasure Island* and boosted the sensitivity to tap into lexical organisation by the sorting task, might still not have had enough power to detect individual participants' differences. If they had, the results would have been more clear-cut in that both distinct individual differences and group differences would have been found. In addition, it is speculated that the differences in L1 and L2 lexical organisations might be subtler than one would expect, and thus would not be easily discerned. This is what Wilks (1999) predicted, and the results this project obtained have supported her claim. It can be argued that such high frequency words as 1K verbs and adjectives might be fully integrated into the L2 mental lexicons and thus they do not bring about substantial L1 and L2 differences through sorting tasks.

Second, it should be noted that the dispersion of individual participants' distance matrices (dendrograms) for both the NS and NNS groups were rather wide, judging from the *SD*s for each group (see Table 7.7). However, when the difference of group distance matrices between the two groups was analysed, the bulk of dispersion of individual participants' results decreased as a result of putting individual results together into a single group of data. Eventually, the NS dendrogram turned out to be statistically different from the NNS dendrogram. In this regard, veiled individual differences between L1 and L2 participants became evident by applying a permutation test to the overall data as reported in the previous section.

Third, in cluster analysis, distances of inter-object relationships in an individual participant's dendrogram (distance matrix) are established as being represented by binary values of either "zero" or "one". In cluster analysis of the sorting task results, a value of "zero" represents a perfect similarity between lexical items (e.g., if a participant puts *blue, black* and *red* into a cluster, then the similarity (distance) between these words is "zero"). Meanwhile, a value of "one" represents a perfect dissimilarity between lexical items (e.g., if a participant puts *happy* in a cluster, *red* in another cluster, and *near* in yet another cluster, then the similarity (distance) between these three lexical items is "one", or in other words perfectly dissimilar.). This binary-based data structure of the individual sorting task results made it extremely difficult to detect individual differences when individual results were examined solely by themselves. However, when these individual results were put together for further analysis by means of cluster analysis and a permutation test, then the underlying individual differences became evident. Regarding the co-occurrence (distance) matrix that shows the number of times words were associated with other words produced by participants, the value in the C_{jk} cell shows the number of participants who put word$_j$ and word$_k$ in the same cluster and reveals the degree of similarity of C_{jk} participants had (Appendices 7.3a and 7.3b). As Deese (1965) and Preece (1976) argued, meaning and its structure in memory is not confined to individuals, but rather a property of a culture. Also, group data analysis is not only related to group data but also individual participants. Preece states that "pooling individual data will produce more valid results" (p. 5) when he analysed psycholinguistic data produced by word association tests and a tree-construction test. These claims also hold true with the present data analysis, and thus the results of a permutation test run on the matrices of the two groups revealed not only group differences but also individual differences in the sorting task results.

Thus, it is concluded that the present experiment using 1K adjectives confirmed the individual differences in lexical organisation as well as group differences between the NS and NNS groups.

7.4.3 Evidence for organisational differences between L1 and L2 mental lexicons

This section addresses two distinctive differences in L1 and L2 organisations that the present sorting task uncovered. They are the L1 and L2 differences in (a) cluster number and composition and (b) participants' sorting behaviour of polysemous words.

As reported in 7.3.4, the NS group made five final clusters of lexical items, whereas the NNS group made three. The NS' final clusters were comprised of the "Positive meanings and emotions" cluster, the "Negative meanings and emotions" cluster, the "Physical dimension of degrees" cluster, the "Physical dimension of time, space, shape and size" cluster, and the "Physical dimension of colours" cluster. It is noted that L1 participants clearly distinguished positive meanings and emotions from negative meanings and emotions. They also differentiated adjectives of physical dimension into three distinctive clusters: the physical dimension of degrees, the physical dimension of time, space, shape and size, and the physical dimension of colours. Thus, L1 lexical organisation was composed of five distinctive sets of final clusters. Meanwhile, the NNS' final clusters were made up of the "Positive and negative meanings and emotions" cluster, the "Physical dimension of time, space, degrees, shape and size" cluster, and the "Physical dimension of colours" cluster. Thus, L2 participants did not distinguish positive meanings and emotions from negative ones, but rather they merged them into a single cluster of positive and negative meanings and emotions. This suggests that in L2 lexical organisation there are some undeveloped areas which are not distinctively separated from each other, while in L1 organisation they are clearly separated from each other. This also holds for the case with adjectives of physical dimension. The NS group made three different final clusters of physical dimension, but the NNS group made only two final clusters of physical dimension. Among them, both groups shared the "Physical dimension of colours" (*black, blue, red*), indicating that the colours dimension is an obvious, distinct cluster that is firmly integrated into the lexical organisations of both participant groups. Thus, besides the colours dimension, the NNS group merged all adjectives of the physical dimension into a single big cluster of the "Physical dimension of time, space, degrees, shape and size" cluster. These results show that L2 lexical organisation is not developed to such an extent that it is distinctively separated from each other in semantically-related neighbouring clusters as L1 lexical organisation is. It should be noted that the sorting task using 1K verbs (Chapter 6) produced the same underlying tendency of the participants in that L1 participants made a larger number (eight) of clusters than their L2 counterparts' (six).

Such high frequency words as the 1K adjectives used in the present experiment are usually polysemous and this fact might have affected the results. That is, L2 participants in the present sorting task appear to have accessed the most common meanings in their mental lexicons of the tested adjectives, which were often different from those in the native speakers' mental lexicons. Eventually, the sorted results were different from each other between the two groups. Table 7.12 summarises the polysemous adjectives that L1 and L2 participants put into different final clusters or linked to a different word or cluster.

Table 7.12. Cluster analysis results comparison: Polysemous words that L1 and L2 participants sorted differently

Word	NS		NNS	
	Strongly linked word/cluster	Final cluster	Strongly linked word/cluster	Final cluster
open	the *natural* cluster	Positive	the *near* cluster	$Physical_3$
strong	the *nice* cluster	Positive	*hard*	$Physical_3$
wrong	*right*	Positive	*bad*	Nega & Posi
right	*wrong*	Positive	the *true* cluster	Nega & Posi
single	*alone*	Negative	*alone*	$Physical_3$
alone	*single*	Negative	*single*	$Physical_3$
hot	the *light* cluster	$Physical_1$	the *strong* cluster	$Physical_3$
light	*hard*	$Physical_1$	the *little* cluster	$Physical_3$
hard	*light*	$Physical_1$	*strong*	$Physical_3$
low	*high*	$Physical_2$	the *long* cluster	$Physical_3$
high	*low*	$Physical_2$	*wide*	$Physical_3$
round	the *wide* cluster	$Physical_2$	the *low* cluster	$Physical_3$

Note. A strongly linked word or cluster was identified between the word in question and the word or cluster which was most closely related to the word in question in dendrograms (Figures 7.1 and 7.2). Positive = Positive meanings and emotions; Negative = Negative meanings and emotions; Nega & Posi = Positive and negative meanings and emotions; $Physical_1$ = Physical dimension of degrees; $Physical_2$ = Physical dimension of time, space, shape and size; $Physical_3$ = Physical dimension of time, space, degrees, shape and size.

Table 7.12 shows that L1 participants sorted polysemous words into clusters in a different way from the way L2 participants did in many respects. L1 participants linked *open* with the *natural* cluster (*natural, clear*) and *strong* with the *nice* cluster (*nice, kind, happy,* etc.) in the "Positive meanings and emotions" cluster. Meanwhile, L2 participants linked *open* with the *near* cluster *(near, far)* and *strong* with *hard* in the "Physical dimension of time, space, degrees, shape and size" cluster. These L1 and L2 differences in sorting polysemous words were also observed in their sorting of *single* and *alone*. While L1 participants linked *single* and *alone* in the "Negative meanings and emotions" cluster, L2 participants linked *single* with *alone* in the "Physical dimension of time, space, degrees, shape and size" cluster. Moreover, for L1 participants, *wrong* and *right* were a set of distinctive antonyms, but, for L2 participants, *wrong* was linked with *bad*, being a kind of synonymous word with it and *right* with the *true* cluster (*true, sure, certain*). Also, L1 participants made a separate final cluster of the "Physical dimension of degrees" with *hot, light* and *hard*, while L2 participants did not make this distinctive cluster but linked *hot* with the *strong* cluster (*strong, hard*), *light* with the *little* cluster (*little, small,* etc.), and *hard* with *strong*. L1 participants made a distinct antonym linkage between *high* and *low* in the "Physical dimension of time, space, shape and size" cluster, whereas L2 participants linked *low* with the *long* cluster (*long, wide, high, deep*) and *high* with *wide*.

Given the above results as a whole, L2 participants' lexical organisation of 1K adjectives was clearly different from its L1 counterpart, particularly regarding contrastive pairs. Contrastive pairs (e.g., *big-little, bad-good, tall-short, and high-low*) are ones of the most important semantic relations among adjectives (Deese, 1964, 1965, Miller & Fellbaum, 1991). The present results indicate that L2 English speakers have developed different structures of contrastive pairs of adjectives in lexical organisation from those L1 English speakers have.

Thus, by investigating the sorting behaviour of NS and NNS for high frequency adjectives, it was found that the two groups are distinctively different from each other in many regards, suggesting that lexical items (1K adjectives) in L2 mental lexicons are linked and structured markedly differently from those of L1 mental lexicons.

7.5 Conclusion

The present chapter revealed some distinct differences in the sorting behaviour of the NS and NNS groups as well as their lexical organisations that were tapped into by the sorting task. While on average L1 participants completed the task almost as quickly as their L2 counterparts as in the previous experiment, they took statistically significantly more time in completing the sorting task using 1K adjectives than the task using 1K verbs. Thus, L1 English speakers appear to be more sensitive to the types of words used in the sorting tasks than L2 English speakers as shown by the larger difference in processing time between tasks. L1 participants made a larger number of clusters and a smaller cluster size than their L2 counterparts, but the differences were not statistically significant. This project has not yet produced any distinct results on this issue.

Introduction of permutation tests during data analysis confirmed interesting findings of underlying L1 and L2 differences in the sorting task results. Although L1 participants were less varied than their L2 counterparts regarding their degree of goodness-of-fit against the group data, the difference was not statistically significant. Also, a statistically significant difference was not found regarding the distances of individual participant's dendrograms between the two groups. However, the two group dendrograms were statistically significantly different from each other. Meaning and its organisation in cognitive structure is not confined to individuals, but rather a property of a culture (Deese, 1965; Preece, 1976). Considering group data analysis performed on the results of psycholinguistic experiments should be not only related to group data but also individual participants, the present sorting task revealed individual differences in lexical organisation as well as group differences between the NS and NNS groups. Furthermore, analysis of organisational differences between L1 and L2 mental lexicons showed that the two structures are distinctively different from each other in important ways including the number of final clusters, composition, and contrastive pairs of adjectives. Thus, it appears to be the case that the two lexical organisations have the same degree of lexical density at least in regard to high frequency lexical items (1K verbs and nouns) but they are differently structured.

In Chapter 8, which reports the results of the last sorting task using 1K nouns, these hypotheses will be tested while also addressing the issue of whether the hypotheses can be confirmed as L1 and L2 differences in lexical organisation.

Chapter 8: Predictors of L1 and L2 differences in sorting results

8.1. Introduction

This chapter reports the final sorting task experiment of the present research project, which addresses four main research questions.

First, I will address whether the phenomenon confirmed in Chapters 6 and 7, that the NS and NNS groups do not differ in time length to complete a sorting task, can be generalised irrespective of word types used in the sorting tasks (research question 1a). Moreover, in Chapter 7, a comparison of verb sorting and adjective sorting task results revealed that there was a word type effect on task completion time. L1 participants took significantly more time in completing adjective sorting than verb sorting, whereas L2 participants did not show such a clear difference in task completion time. This will be further addressed using 1K nouns in the present sorting task (research question 1b).

Second, the central theme in this chapter is to make a final attempt at finding predictors of L1 and L2 differences in sorting task results at the individual participant level. A main question that will be addressed is whether nouns, another major word class that has not yet been tested in the present project, can reveal that L1 participants are different from their L2 counterparts in the number, size and variability of clusters they make (research question 2), as Meara and Schur (2002) have reported. That is, can nouns produce L1 and L2 differences of cluster structures in lexical organisation or not? Do L1 participants perceive words by means of distinct sets and produce a larger number of clusters and a smaller cluster size, with the results being less varied, than their L2 counterparts? In L1 lexical acquisition, nouns are usually the earliest to be learned (Gentner, 1981). Moreover, nouns are usually easier to learn than other word classes. Pease, Gleason and Pan (1993) state that the "preponderance of common nouns (or general nominals) in early vocabularies has been noted in many studies. Even though a particular noun may be used less frequently than a particular verb in the language, nouns appear to be acquired more easily and utilized when children are at a loss to remember a particular verb" (p. 123). Similarly, in L2 lexical acquisition, nouns are usually learned predominantly in the early stages (Källkvist, 1998, 1999). Furthermore, Dietrich (1989) reports that "the very early lexicon of the adult learner, like that of the child, is mainly a nominal one. The more or less rapid development of the verbal category is a subsequent process. The overall picture, then, is the same as in first language acquisition" (p. 19). This nature of nouns "being the earliest and the easiest" in both L1 and L2 lexical acquisition might mean that nouns have established some undiscovered unique structures in the mental lexicon that were not detected in the verbs and adjectives used in this project.

Third, in relation to the variability issue addressed in research question 2, this chapter explores whether L1 lexical organisation is less varied than their L2 counterparts, when it is tested by the degree of goodness-of-fit of the individual results against the group data by means of a permutation test (research question 3). In Chapter 7, a permutation test failed to detect a difference. If the present experiment still fails to confirm a difference in L1 and L2 variability, it would lead to the conclusion that the difference of the two (i.e., L1 and L2) lexical organisations are more subtle than one would expect and support Wilks's (1999) prediction. If some substantial differences between the NS and NNS groups are found, it would mean that the degree of similarity between the L1 and L2 lexical organisations varies depending on which word type (e.g., nouns, verbs and adjectives) an experiment is using. This final experiment will reach a conclusion on this issue.

Lastly, I will answer the question of whether L2 lexical organisation is differently arranged than its L1 counterpart (research question 4). This is concerned with Wilks and Meara's (2002) prediction on the difference in lexical organisation, which was supported by the previous sorting task results using 1K verbs (Chapter 6) and 1K adjectives (Chapter 7). This chapter will confirm whether the same results were obtained by the sorting task using 1K nouns and if the prediction Wilks and Meara made can be generalised.

The above-mentioned points are summarised below:

1.a Do L2 participants complete the sorting task as quickly as L1 participants?
1.b Do participants need to take a different amount of time to complete sorting tasks depending on the word classes of the words used in them?
2. Do L1 participants make a larger number of clusters and fewer words per cluster than their L2 counterparts do while the results L1 participants produce are less varied than their L2 counterparts?
3. Is the L2 lexical organisation less consistent than the L1's, where the sorting task results of L2 participants are more varied than their L1 counterparts?
4. Is the L2 lexical organisation structurally different from its L1 counterpart?

8.2 Method

8.2.1 Participants

As was the case with previous experiments, there were two participant groups in this experiment. The first was comprised of 30 adult, native speakers of English (NS). They were teachers of English in Kumamoto, visiting scholars and students studying at Kumamoto University. The second consisted of 30 adult, advanced-level Japanese speakers of English (NNS). They were either English teachers at the college level or persons having a high competence of English as judged by a TOEFL score of 213 or more on the computer-based version or a score of 550 or more on the paper-based version or a TOEIC score of 730 or more that had been taken within the last two years.

8.2.2 Data Collection

Fifty nouns were used in the present sorting task. As was the case with the verb (Chapter 6) and adjective (Chapter 7) sorting tasks, 50 nouns were randomly selected from among all the nouns contained in Robert Stevenson's (1883), *Treasure Island*, "Chapter 32. The Treasure-hunt — The Voice Among the Trees". The 50 nouns (which were among the first, 1000 high frequency words from the JACET 8000) selected for the experiment were as follows: *air, blood, child, colour, effect, effort, eye, face, fear, field, foot, front, ground, hand, head, influence, island, life, line, man, middle, mind, mistake, mouth, note, part, party, play, power, rest, sea, sense, shoulder, sign, son, song, sound, spirit, star, talk, thought, top, train, tree, view, voice, west, wonder, wood, word.*

After piloting, the task was administered to each of the participants individually, where the directions were identical with the ones used in Chapters 6 and 7 (except for the type of words tested).

8.2.3 Data Analysis

To answer research questions 1a, 1b and 2, the sorting task data was analysed by comparing the means and *SD*s of the L1 and L2 results as was done in the previous experiments. This was performed by means of a *t*-test and an *F*-test on the time taken to complete the task, cluster number and size and the "largest" cluster participants made. Also, in regard to a within-group comparison of the amount of time taken to complete the three sorting tasks (i.e., verbs, adjectives and nouns), an analysis of variance (ANOVA) was run on the NS and NNS group data. Then, multiple comparisons using Bonferroni tests were made to look for a pair of task results that had a statistically significant difference in task completion time, if any.

To address research question 3 of whether the sorting task results of participants of the NNS group were more varied than those of the NS group, a permutation test of individual differences between groups was conducted. A permutation test in the present analysis made use of the distance matrices of individual participants, where a distance matrix is the sum of the lengths of each participant's dendrogram tree. Then each individual participant's distance matrix data was examined to find its degree of goodness-of-fit with the (L1 or L2) group dendrogram.

To answer research question 4, three types of analyses were made. First, while addressing the individual differences in the sorting task results of the NS and NNS groups, a *t*-test was run on individual participants' distance matrices. Second, each of the two group dendrograms was depicted as a final graphic representation of cluster analysis and their components (i.e., lexical items) were compared to each other to analyse the qualitative differences in lexical organisation. Third, a permutation test was run on the group data to answer the question of whether the L1 and L2 group dendrograms were statistically significantly different from each other.

8.3 Results

8.3.1 Time taken to complete task

In regard to research question 1a, Table 8.1 shows the means and *SD*s of the amount of time the NS and NNS groups took to complete the sorting task using the 50 high frequency nouns selected from *Treasure Island.*

Table 8.1. Time to complete 1K *Treasure Island* nouns sorting task

	NS (n = 30)	NNS (n = 30)	*t*-value	*F*-value
Mean	8.41	8.60	0.24n.s.	0.65n.s.
SD	3.31	2.67		

Note. n.s. = not significant.

On average, NNS took 0.19 minutes (11 seconds) more to complete the task than NS did. An unmatched *t*-test revealed that there was no statistically significant difference between the two means ($t(58) = 0.24$, n.s.). L2 participants completed the sorting task almost as quickly as their L1 counterparts. This gives further evidence to the finding made in Chapters 6 and 7, namely that the way in which L1 and L2 participants carried out the sorting tasks was not substantially different from each other, irrespective of the word types used in them.

In regard to question 1b, Table 8.2 shows the amount of time taken for NS to complete the verb, adjective and noun sorting tasks. The results revealed that NS completed the verb sorting the quickest followed by the noun and adjective sorting tasks.

Table 8.2. NS' time taken to complete the sorting tasks

	verbs (n = 30)	adjectives (n = 30)	nouns (n = 30)
Mean	7.94	9.96	8.41
SD	3.28	3.31	3.31

Note. n.s. = not significant.

Table 8.3. Repeated measures ANOVA: Word class and NS task completion time

Sources of variance	Sums of Square	*df*	Mean Square	*F*
Word class	66.94	2	33.47	3.07n.s.
Error	949.77	87	10.92	
Total	1016.77	89		

Note. n.s. = not significant.

Table 8.3 shows the results of a Repeated Measures ANOVA run on the data in Table 8.2. Regarding the NS results, a word class effect was not detected between the three sorting tasks at the 5% significance level ($p = .0517$). Thus, the *p*-value failed to reach significance by only a slim margin. In Chapter 7, which addressed the word type effect on the sorting task results, a wide difference in task completion time between verb sorting (= 7.94 min.) and adjective sorting (= 9.96 min.) was revealed. The 2.02 minutes' difference was statistically significant ($t(58) = 2.37, p < 0.05$). Accordingly, to examine whether there was any substantial difference in completion time between the three sorting tasks for the NS group, multiple comparisons using a Bonferroni test was run. Table 8.4 shows the results.

Table 8.4. Multiple comparisons of task completion time: Bonferroni test results of the NS sorting tasks

Compared task pair	Mean difference (min.)	Standard error	*p*-value
Verbs vs. Adjectives	2.02n.s.	0.94	0.12
Verbs vs. Nouns	0.47n.s.	0.95	1.00
Adjectives vs. Nouns	1.55*	0.53	0.02

Note. Verbs = Verb sorting task; Adjectives = Adjective sorting task; Nouns = Noun sorting task; $*p < 0.05$, n.s. = not significant.

Table 8.4 reveals that the mean difference in task completion time was statistically significant between the adjective and noun sorting tasks for the NS group ($\alpha = 5\%$), but there was no statistically significant difference between the other pairs. It should be noted that in the NS results, the mean difference between the verb sorting and adjective sorting tasks was the largest, but the difference was not statistically significant. This is because the large standard error (= 0.94) had an impact on the measurement. That is, the mean difference between the verb and adjective sorting tasks was substantial, but the dispersion of the results was large from the predicted (linear) regression line and thus no reliable data was obtained.

Table 8.5 shows the amount of time NNS took to complete the three sorting tasks. It was found that NNS completed the noun sorting task the quickest, followed by the verb and adjective sorting tasks. Sorting adjectives took both the NS and NNS groups the longest amount of time among the three sorting tasks. It should be noted that the NNS's dispersion in task completion time was the largest for the verb sorting task.

Table 8.5. NNS' time taken to complete the sorting tasks

	verbs ($n = 30$)	adjectives ($n = 30$)	nouns ($n = 30$)
Mean	9.55	10.64	8.60
SD	3.89	3.32	2.67

Table 8.6 shows the results of a Repeated Measures ANOVA run on the data in Table 8.5. A word class effect was not detected at the 5% significance level (p =.065).

Table 8.6. Repeated measures ANOVA: Word class and NNS task completion time

Sources of variance	Sums of Square	*df*	Mean Square	*F*
Word class	62.73	2	31.36	2.83n.s.
Error	964.62	87	11.09	
Total	1027.35	89		

Note. n.s. = not significant.

As was the case with the NS results, multiple comparisons by means of a Bonferroni test were run to examine whether there was any substantial difference in completion time between the three sorting tasks for the NNS group. Table 8.7 tabulates the results.

Table 8.7. Multiple comparisons of task completion time: Bonferroni test results of the NNS sorting tasks

Compared task pair	Mean difference (min.)	Standard error	p-value
Verbs vs. Adjectives	1.09n.s.	1.01	0.86
Verbs vs. Nouns	0.95n.s.	0.87	0.86
Adjectives vs. Nouns	2.04*	0.38	0.00

Note. Verbs = Verb sorting task; Adjectives = Adjective sorting task; Nouns = Noun sorting task; *$p <$ 0.05, n.s. = not significant.

Table 8.7 reveals that the differences in mean task completion time were only statistically significant between the adjective and noun sorting tasks for the NNS group ($\alpha = 5\%$). However, there was no statistically significant difference in the means of other pairs. This is exactly the same result as that of the NS group.

Given these results as a whole, it was found that there were significant word type (i.e., word class) effects in task completion time for both the NS and NNS groups.

8.3.2 Cluster number, size and variability

To answer to research question 2, comparisons were made on the mean number of clusters, mean number of words per cluster and the mean largest cluster the NS and NNS groups made.

Table 8.8 shows the mean number of clusters participants made in the present sorting task, where the count excludes single, isolated words. The table reveals that on average NS produced a larger mean cluster number than NNS did, and the difference was statistically significant (t (58) = 2.65, $p < 0.01$.). As the *SD* values show, the variance of the NS was slightly larger than that of the NNS. However, an F-test showed that there was no meaningful difference in variability between the two groups ($F(29) = 0.86$, n.s.). Thus, it was revealed that nouns brought about substantial L1 and L2 differences in cluster number in a sorting task, whereas verbs and adjectives had failed to do so.

Table 8.8. Mean number of clusters (which excludes single, isolated words)

	NS ($n = 30$)	NNS ($n = 30$)	t-value	F-value
Mean	8.93	7.47	2.65**	0.86n.s.
SD	2.23	2.06		

Note. **$p < 0.01$, n.s. = not significant.

In answer to the question of whether the L1 participants would make a smaller amount of words per cluster than their L2 counterparts, the relevant data was tabulated and is shown in Table 8.9.

Table 8.9. Mean number of words per cluster (which excludes single, isolated words)

	NS ($n = 30$)	NNS ($n = 30$)	t-value	F-value
Mean	5.61	6.61	2.12*	1.28n.s.
SD	1.69	1.91		

Note. *$p < 0.05$, n.s. = not significant.

Table 8.9 reveals that on average NS produced a smaller mean number of words per cluster than NNS did, and the difference was statistically significant (t (58) = 2.12, $p < 0.05$.). As the *SD*s show, the variance of the NS group was slightly smaller than its NNS counterpart. But an F-test revealed that there was no substantial difference in variability between the two groups ($F(29) = 1.28$, n.s.). An inverse relationship had been predicted to exist in that if the mean number of clusters the two groups made were significantly different from each other, directly related data of the mean number of words per cluster would produce a similar difference. The present sorting task generated the difference as predicted.

Furthermore, Table 8.10 shows that on average L2 participants produced a bigger mean "largest" cluster than their L1 counterparts, and the difference was statistically significant ($t(58) = 2.32, p < 0.05$.). The results suggest that L2 participants have not developed L2 words belonging to distinct sets as much as L1 participants have. As the *SD*s show, the variance of the NS group was slightly larger than its NNS counterpart. But an *F*-test revealed that there was no substantial difference in variability between the two groups ($F(29) = 0.81$, n.s.).

Table 8.10. Mean largest cluster participants made

	NS ($n = 30$)	NNS ($n = 30$)	*t*-value	*F*-value
Mean	10.13	12.07	2.32*	0.81n.s.
SD	3.39	3.05		

Note. $*p < 0.05$, n.s. = not significant.

Putting these results together, it was found that although there was no substantial difference in variability, nouns revealed distinct, statistically significant differences between the NS and NNS groups in cluster number, size and "largest" cluster made. L2 English speakers, even if they are at an advanced level, have not yet developed native-like lexical organisation. Nouns appear to be the most evident predictors of this discrepancy.

8.3.3 Comparison of L1 and L2 dendrograms: Individual dispersion of goodness-of-fit with group data

This section will answer research question 3 of whether the L2 lexical organisation is less consistent than L1's and if the sorting task results of L2 participants are more varied than their L1 counterparts. For the investigation, I computed the degree of goodness-of-fit of individual NS participants' dendrograms against the group's dendrogram as calculated by the square of the Minkowski distance between the distance matrices (= dT_NS), and the degree of goodness-of-fit of individual NNS (Japanese) participants' dendrograms against the group's dendrogram as calculated in the same way (= dT_JP). See Appendix 8.1 for the tabulated results of each calculation. Table 8.11 shows the T values (i.e., average dispersion of dendrograms of individual participants against the dendrogram of the NS and NNS results, respectively), where the smaller the value is, the better the individual results fit with the group result. In the table, the *p*-value shows whether the difference between the results of the NS and NNS is statistically significant. The T value of NS fitted better with the group result than that of NNS, suggesting that NS results were less varied than NNS's results. The difference examined by a permutation test reached the 5% significance level ($p = 0.017$). Thus, it was found that the NNS individual results were more varied than their NS counterparts when 1K nouns were used in a sorting task. This provides evidence for previous word association test-oriented studies (e.g., Fitzpatrick, 2006; Meara, 1978, 1983; Meara & Schur, 2002; Postman & Keppel, 1970; Riegel & Zivian, 1972; Szalay & Deese, 1978) which reported that NS results are less varied than NNS results are.

Table 8.11. Permutation test result: Average dispersion of dendrograms of individual participants against dendrogram of the group data

	NS ($n = 30$)	NNS ($n = 30$)	p-value
T	71.48	95.92	0.017*

Note. T = Average dispersion of dendrograms of individual participants against the dendrogram of the group data; p-value = significant level of the difference of the T value between the NS and NNS results as calculated by a permutation test which was administered on the assumption that the average dispersion of the NS group's individual dendrograms and the NNS's would be equal to each other; $*p < 0.05$.

8.3.4 Organisational differences of L1 and L2 cluster structures

This section will answer research question 4 addressing whether the L2 lexical organisation is structurally different from its L1 counterpart, which is something that should be unearthed in the sorting task results. Analysis was done in two ways. First, while addressing the issue of individual differences in L1 and L2 sorting task results (where single, isolated words were excluded), a t-test was run on the individual participants' distance matrices for both groups, where a distance matrix is the sum of the length of dendrogram trees (see Appendix 8.2). Table 8.12 shows the means and *SD*s of the distance of each individual participant's dendrogram (distance matrix) for both the NS and NNS groups.

Table 8.12. Distance of each individual participant's dendrogram

	NS ($n = 30$)	NNS ($n = 30$)	t-value	F-value
Mean	7.93	6.47	2.65**	0.86n.s.
SD	2.23	2.06		

Note. $**p < 0.01$; n.s. = not significant.

Table 8.12 reveals that the distance of the NNS group was smaller than the NS group, suggesting that on average L2 participants made clusters having a shorter distance than their L1 counterparts. An unmatched t-test revealed that there was a statistically significant difference between the two means ($t(58) = 2.65, p < 0.01$.). Thus, it was confirmed that the

NNS group made a substantially smaller number of clusters than the NS group in the noun sorting task.

Second, cluster analyses were run on the matrices of the NS and NNS group data (Appendices 8.3a and 8.3b). Figure 8.1 is the dendrogram of the results computed for the NS group and Figure 8.2 is the dendrogram of the results for the NNS group.

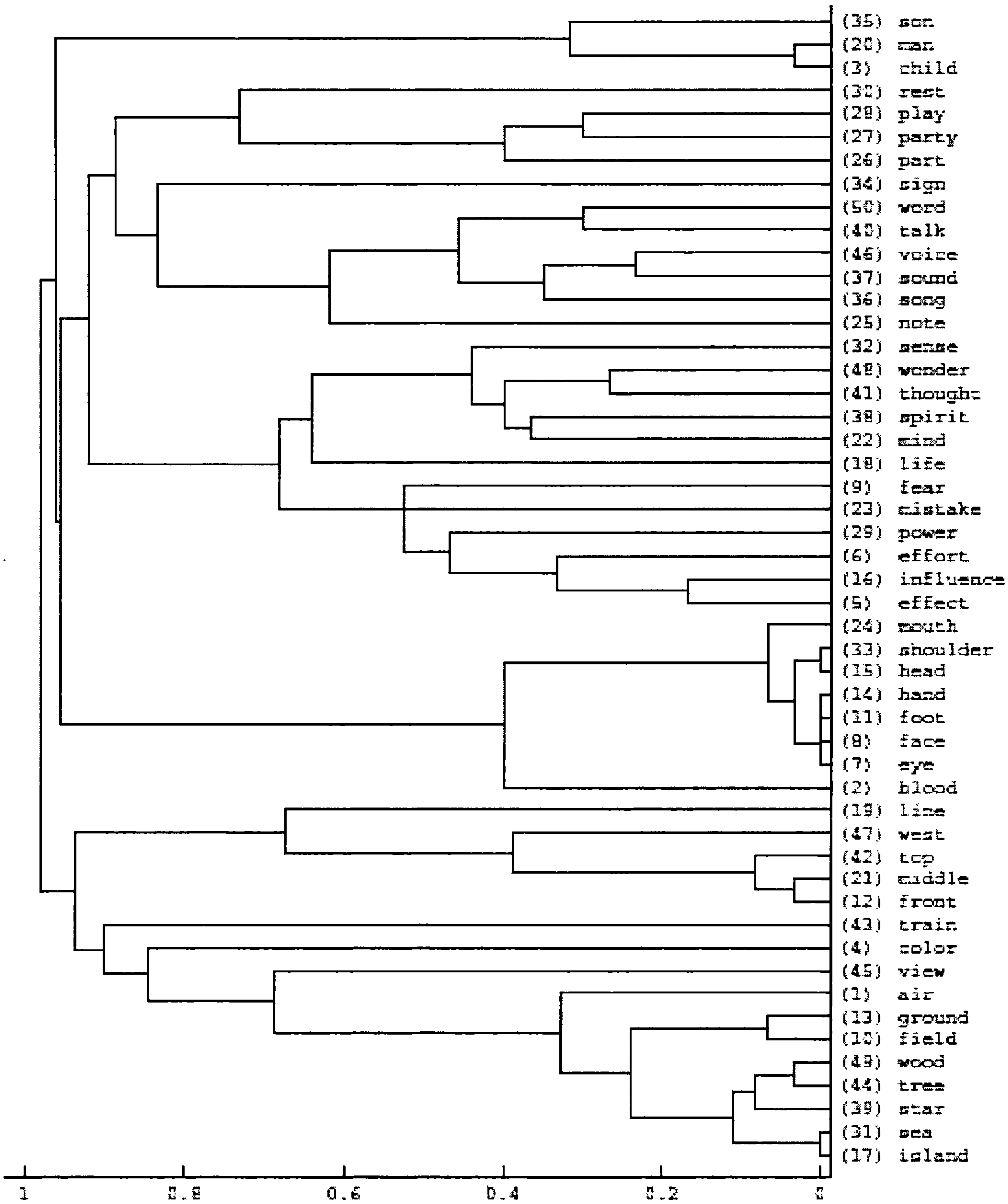

Figure 8.1. Dendrogram of NS's sorting task results. Analysis was done by means of cluster analysis. Numbers in parentheses show the numerical order of the word in the co-occurrence matrix of Appendix 8.3a.

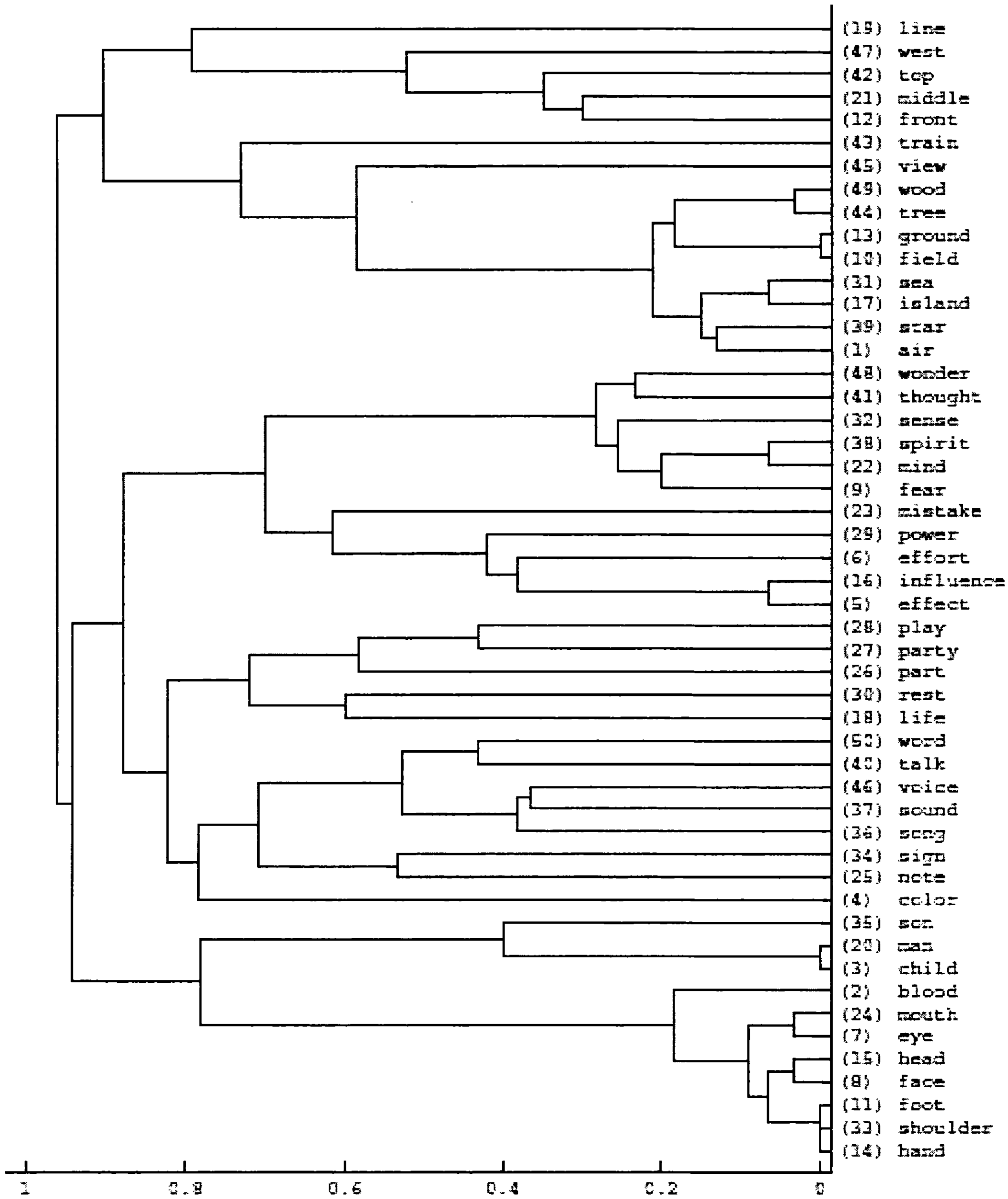

Figure 8.2. Dendrogram of NNS's sorting task results. Analysis was done by means of cluster analysis. Numbers in parentheses show the numerical order of the word in the co-occurrence matrix of Appendix 8.3b.

The number of clusters was finalised based on the dendrograms shown in Figures 8.1 and 8.2. As was done in Chapters 6 and 7, the identification of the number of final clusters of a dendrogram was carried out by following stopping rules. As a result, a set of six final clusters were located in the L1 dendrogram, whereas a set of five final clusters were identified in the

L2 dendrogram. This result was consistent with the previous sorting task results for 1K verbs (Chapter 6) and 1K adjectives (Chapter 7) in that the NS group made a larger number of clusters than the NNS group did in each case. Thus, it was substantiated that L1 participants made a larger number of clusters than their L2 counterparts, irrespective of the type of word class being tested in a sorting task.

Final clusters identified in the NS and NNS dendrograms are shown in Tables 8.13 and 8.14, respectively. The final clusters for the NS group were the "Dimension of activities and communication" cluster, the "Dimension of cognition and effect" cluster, the "Dimension of human/animal body" cluster, the "Dimension of geometries" cluster, the "Train" cluster, and the "Dimension of elements in nature" cluster. Meanwhile, the final clusters for the NNS group were the "Dimension of activities and communication" cluster, the "Dimension of cognition and effect" cluster, the "Dimension of man and human/animal body" cluster, the "Dimension of geometries" cluster, and the "Dimension of *train* and elements in nature" cluster. It should be noted that both groups produced the exact same "Dimension of geometries" (*line, west, top, middle, front*) cluster. For the rest, the final clusters of the two groups were different from each other in both number and composition. This lends support to the claim that there are structural differences between L1 and L2 lexical organisations, which the previous sorting task using the 1K adjectives also affirmed. Thus, the prediction that Wilks and Meara (2002) made about L1 and L2 lexical organisations having the same lexical density but different structures was substantiated.

Table 8.13. Final clusters identified for the NS group ($k = 50$)

No.	Cluster	Word No.	Words
1	Dimension of activities and communication	14	*son, man, child, rest, play, party, part, sign, word, talk, voice, sound, song, note*
2	Dimension of cognition and effect	12	*sense, wonder, thought, spirit, mind, life, fear, mistake, power, effort, influence, effect*
3	Dimension of human/animal body	8	*mouth, shoulder, head, hand, foot, face, eye, blood*
4	Dimension of geometries	5	*line, west, top, middle, front*
5	Train	1	*train*
6	Dimension of elements in nature	10	*colour, view, air, ground, field, wood, tree, star, sea, island*

Table 8.14. Final clusters identified for the NNS group ($k = 50$)

No.	Cluster	Word No.	Words
1	Dimension of activities and communication	13	*play, party, part, rest, life, word, talk, voice, sound, song, sign, note, colour*
2	Dimension of cognition and effect	11	*wonder, thought, sense, spirit, mind, fear, mistake, power, effort, influence, effect*
3	Dimension of man and human/animal body	11	*son, man, child, blood, mouth, eye, head, face, foot, shoulder, hand*
4	Dimension of geometries	5	*line, west, top, middle, front*
5	Dimension of *train* and elements in nature	10	*train, view, wood, tree, ground, field, sea, island, star, air*

A permutation test was run on the L1 and L2 group data of distance matrices to test whether the L1 dendrogram represented in Figure 8.1 would be different from the L2 dendrogram represented in Figure 8.2. The test confirmed that the two dendrograms were markedly different from each other at the 5% significance level ($p = 0.0388$). This further substantiates the assertion that L1 lexical organisation is different from L2 lexical organisation.

8.4 Discussion

This chapter has thus far shown that nouns can be predictors of L1 and L2 differences in sorting behaviour and lexical organisation. In particular, the sorting task results revealed that L2 speakers, for the most part, "do not yet perceive vocabulary as belonging to smaller, constrained and strongly connected sets and they tend to associate words in more diverse and less predictable ways than native speakers do" (Meara & Schur, 2002, p. 179). On the other hand, L1 speakers appear to be more conscious of the semantic relationships between words and the way words fall into distinctive groups.

In this section, two questions will be posed. First, the question of how the components of the sorting results are different between the NS and NNS groups will be addressed and an examination of L1 and L2 differences in lexical organisation observed in the sorting task results will be made. Second, the question of which factors of nouns brought about such substantial L1 and L2 differences will be addressed. In answering the second question, a comparison of the verb, adjective and noun sorting task results will be made as well.

8.4.1 Aspects of L1 and L2 differences in lexical organisation

A comparison of the final clusters the NS and NNS groups made, which were shown in Tables 8.13 and 8.14, disclosed three important L1 and L2 differences in lexical organisation. First, the NS group made a distinct one word "Train" cluster, but the NNS group combined *train* into the "Dimension of *train* and elements in nature" cluster. The verb sorting experiment also produced a similar result in that L2 participants were not always aware of lexical items belonging to distinct sets, whereas L1 participants seemed to be. That is, the NS group made a distinctive one word cluster of *die*, but the NNS group merged *die* into the "Dimension of FIGHT" cluster. NNS have developed a cluster structure where they do not always perceive lexical items as distinct sets.

Second, the NS group placed *son, man,* and *child* into the "Dimension of activities and communication" cluster, but the NNS group placed these three words into the "Dimension of man and human/animal body" cluster. Thus, although *son, man* and *child* were sorted into a single cluster by both groups similarly, the structured clusters were not the same, suggesting that the three word cluster is integrated into L1 and L2 lexical organisations in a different way.

Third, the NS group made a distinctive cluster, namely the "Dimension of human/animal body" *(mouth, shoulder, head, hand,* etc.*)* cluster. Meanwhile, the NNS group did not separate this cluster from the words *son, man* and *child* and instead made a merged cluster of the "Dimension of man and human/animal body". Thus, the cluster the NS group made was composed of lexical items that were exclusively concerned with the human/animal body, but this was not the case with the NNS group. These results confirmed that L2 lexical organisation has not yet developed as L1 lexical organisation has to the extent that clusters in the organisation are distinctively separated from each other in light of semantically-related, neighbouring clusters.

A further comparison of the final clusters that the two groups made revealed that L2 participants, even at an advanced-level, have not established L1 English-like semantic structures. This was confirmed by observing pairs (i.e., clusters) of lexical items having the highest similarity in Figures 8.1 and 8.2. For example, NS linked *head* with *shoulder*, whereas NNS connected *head* with *face*. Contrastive linguistics shows that the English word *head* bears a wider range of meanings than the Japanese counterpart (Kunihiro, 1999). In English, *head* is "the part of the body on top of the neck containing the eyes, nose, mouth and brain" (*Oxford Advanced Learners' Dictionary* (7th ed.), 2005). Reflecting this meaning of *head* in the English language and its integration into the L1 English mental lexicon, *head* was on the same hierarchical level with *shoulder* in the NS sorting results. Meanwhile, in the Japanese language, *atama* (*head*) is contrastive with *kao* (*face*), and these two words are separated from each other. This point is reflected in the NNS results where *head* and *face* were on the same taxonomical level and formed a pair. Similar examples were found in the cases of *air* and *view* (NS) vs. *air* and *star* (NNS), *sign* and the "*word, talk,* etc." cluster (NS) vs. *sign* and *note*

(NNS), and *sense* and the "*wonder, thought, spirit, mind*" cluster (NS) vs. *sense* and the "*spirit, mind, fear*" cluster (NNS).

It should be noted that the 1K nouns used in the present experiment were sorted in a different way in the semantic fields of L1 and L2 lexical organisations. Semantic fields are originally concerned with lexical organisations where "the relevant lexical units precisely mark out each other's territory, so to speak" (Singleton, 2000, p. 67). In the present research, semantic fields are defined as final clusters identified as dimensions of grouped lexical items that share some common semantic properties. Examples of semantic fields are human/animal bodies, physical acts, spatial relations and emotions. An examination of the sorting task results for the NS and NNS groups led to the discovery that there were some lexical items which had been sorted differently by the NS and NNS groups. Also, there were occasions when NNS linked a word most closely to a different word from that of NS even if the word was clustered in the identical final cluster in both the NS and NNS results. Table 8.15 tabulates the nouns that L1 and L2 participants sorted into different final clusters or linked to different words even though they were placed in the identical final cluster.

Table 8.15. Cluster analysis results comparison: Nouns that L1 and L2 participants sorted differently

Word	NS		NNS	
	Strongly linked word/cluster	Final cluster	Strongly linked word/cluster	Final cluster
son	the *man* cluster	Dimension_1	the *man* cluster	Dimension_4
man	*child*	Dimension_1	*child*	Dimension_4
child	*man*	Dimension_1	*man*	Dimension_4
rest	the *play* cluster	Dimension_1	*life*	Dimension_1
sign	the *word* cluster	Dimension_1	*note*	Dimension_1
note	the *word* cluster	Dimension_1	*sign*	Dimension_1
sense	the *wonder* cluster	Dimension_2	the *spirit* cluster	Dimension_2
life	the *sense* cluster	Dimension_2	*rest*	Dimension_1
mouth	the *head* cluster	Dimension_3	*eye*	Dimension_4
shoulder	*head*	Dimension_3	*foot/hand*	Dimension_4
head	*shoulder*	Dimension_3	*face*	Dimension_4
hand	*foot/face/eye*	Dimension_3	*foot/shoulder*	Dimension_4
train	the *view* cluster	Train	the *view* cluster	Dimension_6
colour	the *view* cluster	Dimension_6	the *word* cluster	Dimension_1
air	the *ground* cluster	Dimension_6	*star*	Dimension_6
star	*wood/tree*	Dimension_6	*air*	Dimension_6

Note. A strongly linked word/cluster was identified between the word in question and the word or cluster which was the most closely related to the word in question in dendrograms (Figures 8.1 and 8.2). Dimension_1= Dimension of activities and communication; Dimension_2 = Dimension of cognition and effect; Dimension_3 = Dimension of human/animal body; Dimension_4 = Dimension of man and human/animal body; Dimension_5 = Dimension of geometries; Dimension_6 = Dimension of elements in nature; Train = The *Train* cluster.

Table 8.15 reveals that NNS sorted words into the various semantic fields differently from the way NS did. For example, NS sorted *life* into the *sense* cluster in the "Dimension of cognition and effect" cluster, whereas NNS most closely linked *life* to *rest* in the "Dimension of activities and communication" cluster. NS put *mouth* into the *head* cluster in the "Dimension of human/animal body" cluster, but NNS most closely linked *mouth* to *eye* in the "Dimension of man and human/animal body" cluster. Furthermore, it was revealed that NS sorted *colour* into the *view* cluster, whereas NNS sorted *colour* into the *word* cluster. NS sorted *rest* into the *play* cluster, but NNS linked *rest* to *life* most strongly. For NS, *sense* belonged to the *wonder* cluster, but for NNS, *sense* was in the *spirit* cluster. Regarding the "Dimension of elements in nature" cluster, NS linked *star* to *wood* and *tree*, but NNS linked *star* to *air* most closely. Thus, by identifying the sorting task results from the viewpoint of semantic fields, it was found that L1 and L2 lexical organisations were different from each other in crucial areas. These results substantiate that L2 lexical organisation has not yet developed qualitatively to the extent that it has established a native-like lexical structure, in addition to the quantitative differences identified in cluster number, size, the "largest" cluster and cluster variability (last of which was confirmed by a permutation test).

Given these findings, a new question arose: Why were nouns the sole class of words to produce statistically significant L1 and L2 differences in certain areas of the sorting task results? This question will be addressed in the next section.

8.4.2 Word type and its effect on degree of integration into L2 lexical organisation

As mentioned earlier, there is a general consensus among L1 and L2 acquisition researchers that nouns are typically the earliest and easiest types of words in both L1 and L2 language acquisition, and they are expected to be integrated into mental lexicons more firmly than other word types (e.g., Dietrich, 1989; Gentner, 1981, 1985; Källkvist, 1999; Pease, Gleason & Pan, 1993). Thus, among the sorting tasks using different word types, the noun sorting task was expected to yield the most similar results between the NS and NNS groups because of this nature of nouns. But the results turned out to be the opposite, where nouns were the sole word type that generated distinct L1 and L2 differences in certain aspects of the sorting task results. Researchers have agreed that nouns are the earliest and the easiest words for learners, irrespective of whether they learn them as L1 or L2 learners. However, in the case of L2 vocabulary learning, the issue seems to be complicated by the effects of the L1 lexical knowledge that L2 learners have acquired. The factors include the semantic networks L2 learners have already established through the acquisition of their L1 lexicon. In other words, some factors related to L1 lexical knowledge and semantic networks might influence L2 learners when developing their L2 lexical organisation. These factors include conceptual knowledge shared between L1 and L2 mental lexicons, word-specific features (e.g., syllable numbers, ease in pronouncing and writing them), word type effects, etc. Given the results examined in the previous sections as a whole, loan words in Japanese that have been taken from English seem to play a vital role in the case of the L2 participants in the present research project.

Loan words comprise a substantial portion of the Japanese language. They account for 10.1% of spoken Japanese (Inagaki, 1991), and 80.8% of loan words imported into Japanese are from English (Tamamura, 1991). Loan words in Japanese that have been taken from English are learned as part of L1 vocabulary in the form of *Katakana* (a Japanese system of phonograms) words that imitate the original pronunciation of the foreign words in question. However, when they are re-learned as L2 vocabulary, loan words do not always function as facilitators to establish native-like L2 lexical knowledge and organisation. That is, lexical knowledge and semantic fields already built in L1 lexical organisation often hinder L2 lexical knowledge and organisation from developing into native-like knowledge and organisation. This is caused by the fact that Japanese people's knowledge and organisation of loan-word nouns in the Japanese language is so thoroughly integrated in their mental lexicons that it hampers the equivalent L2 English nouns from being newly established in their lexicons. As Källkvist (1999) indicated, "nouns are often 'over-represented' in early [L1] vocabularies" (p. 55), and this nature of L1 nouns persistently affects L2 vocabulary learning. In this regard, it should be noted that when English words are integrated into Japanese, the original grammatical functions including syntax and morphology are usually lost, and only the meanings of the original English words are maintained[1].

Further complications take place owing to the tendency that loan words in Japanese often establish semantic fields that are different from those of the original English words. This deviation often leads Japanese EFL learners to have difficulties to learn the original, correct English meanings. More often than not, Japanese EFL learners are not even aware of the fact that they have fossilised the L1 loan word meanings in their mental lexicons and therefore have ceased to advance in mastering their L2 meanings. Wolter (2006) argues that "L1 lexical knowledge can be both a help and a hindrance when forming L2 connections" (p. 741). English loan words in Japanese are evidence for his claim. English loan words that have first been learned as L1 nouns by Japanese EFL learners often prevent them from establishing native-like L2 lexical knowledge and organisation.

It should be noted that English loan words in Japanese are used predominantly as nouns in Japanese (Ando, 1997). This is because there is a huge gap between the English and Japanese

[1] Needless to say, in the Japanese language there are many English "false friends", which have changed the original meanings of the English words into something totally different.

languages in such areas as grammar, orthography and pronunciation. The two European and Asian languages are substantially different from each other in many respects. Eventually most of the English loan words imported into Japanese are transformed into simplified, nominalised vocabulary. Because of this, English loan words in Japanese cannot be cognates in the strict sense of the word. If loan words had the status of cognates, they would make a positive, explicit cross-linguistic transfer and facilitate the establishment of native-like lexical organisations in L2 mental lexicons. Meara (2006) points out that "the first words learned in an L2 do not have any other L2 words to link to, so they must link to L1 words, and words which are cognates will sometimes form cross language links" (p. 642). These types of cross language links as facilitators in L2 lexical acquisition cannot usually be expected to be established in the case of English loan words in Japanese.

Table 8.16 shows the ratio of loan words among the lexical items used in the verb, adjective and noun sorting tasks in the present research project. The table reveals that a large number of tested words in the noun sorting task (39, 78.0%) are English loan words[2].

Table 8.16. Number and percentage of loan words imported from English into Japanese used in the sorting tasks

	Verb (k =50)	Adjective (k = 50)	Noun (k = 50)
Loan words	2	27	39
%	4.0%	54.0%	78.0%

Note. Loan words taken from English were identified by the *Koojien* [The Encyclopaedia of the Japanese Language] (6th ed.). 2006. Tokyo: Iwanami Shoten.

As shown in Table 8.15 in the previous section, there were 16 nouns that NNS sorted into different final clusters or linked to different lexical items from those of NS. Among them, all but one (= 93.8%), namely *son*, were loan words taken from English (*man, child, rest, sign, note, sense, life, mouth, shoulder, head, hand, train, colour, air, star*). As indicated above, for example, L2 participants most strongly linked *life* to *rest*, while L1 participants sorted *life* into the *sense* cluster (which was comprised of *sense, wonder, thought*, etc.). L2 participants linked *head* to *face*, whereas L1 participants linked *head* to *shoulder*. These examples show that NNS

[2] Table 8.16 shows that the number of loan words in the adjective sorting task was also large (= 54.0%). However, no marked effect of those words was found in the adjective sorting results (see Chapter 7). It is speculated that a larger number of the adjectives used in the sorting task have equivalents with a cognate status in Japanese, although they are English loan words. Thus there are smaller gaps between their L1 and L2 meanings and thus English loan words do not seem to hinder the ability of Japanese L2 learners of English to acquire L2 lexical knowledge and organisation.

maintain and transfer their L1 (Japanese) semantic fields which are qualitatively different from English semantic fields during L2 lexical acquisition. This effect lasts even after they began to learn English as an L2 and the effect of L1 transfer was found in the sorting task using English nouns. In other words, these persistent effects of loan words in L2 lexical acquisition exemplify the phenomenon that "the lexical-level connections established during early stages of acquisition may still continue to function under some circumstances once individuals become fluent bilinguals" (Kroll & Tokowicz, 2001, p. 52). Thus, the present sorting task revealed that the most easily and earliest learned words can be barriers for L2 learners in establishing native-like L2 lexical organisation, as seems to be the case with English nouns that are first learned as loan words in Japanese.

It should be noted that verbs have a greater breadth of meanings than nouns and other word classes do (Gentner, 1981, 1985) and that "verb meaning consists of complex and cross-linguistically variable conflation patterns of meaning components" (Källkvist, 1999, p. 54). Therefore, L2 learners encounter and must overcome more challenges in learning L2 verbs than in learning L2 nouns and other word types. This should be an advantage to L2 learners in most cases in L2 lexical acquisition. However, this is not always the case, as shown by English loan words in Japanese, where Japanese L2 learners of English have easily learned them at the early stages of L1 acquisition and integrated and "over-represented" them into their L1 lexical knowledge. Because of their strength and persistence in L1 lexical knowledge and organisation, loan words are stabilised in the L1 mental lexicons of Japanese learners of English. Thus, they can be the hardest L2 lexical items to learn and restructure into native-like L2 knowledge and organisation.

8.5 Conclusion

This chapter addressed the issue of whether nouns can produce distinctive L1 and L2 differences in certain aspects of the sorting task results, which the previous tasks using verbs and adjectives had failed to do. The present experiment confirmed that nouns can be predictors of differences at both the individual participant and group levels. L1 participants produced a larger number of clusters and fewer words per cluster, and their clusters were less varied than their L2 counterparts when examined by a permutation test. Also, the NS and NNS groups produced clusters that were qualitatively different from each other in important respects. An examination of the data revealed that English loan words in Japanese played a role. These words often have different semantic fields from those of English, and they have been firmly cemented in the mental lexicons of Japanese L2 learners of English. Thus, loan words are the most difficult words for Japanese EFL learners to develop native-like lexical knowledge and organisation. A comparison of the verb, adjective and noun sorting tasks with regard to task completion time showed that both groups were sensitive to word types in sorting words and that word types affected the task completion time of both the NS and NNS groups.

This chapter brings to a close the results of the experiments addressing differences in NS and NNS sorting behaviour and results regarding cluster structures of L1 and L2 mental lexicons. In Chapter 9, some of the issues that are central to the project and that have been newly discovered through the experiments will be addressed.

Chapter 9: General Discussion

9.1 Introduction

This thesis project started off with the aim of increasing our understanding of L1 and L2 differences of lexical organisation in the mental lexicon. For this purpose, a series of experiments using sorting tasks were carried out while in particular attempting to reveal the clustering behaviour and underlying lexical structures of L1 and L2 participants. Throughout the endeavour, revisions on the tasks were made so that the revised ones would be better able to examine the L1 and L2 differences and to find out what type of lexical items would be the predictors of the differences. As reported in Chapter 8, we found that nouns, which were selected from a cohesive *Treasure Island* passage, can be the best predictors of L1 and L2 differences in lexical organisation when we gave the NS and NNS groups the sorting task under the direction of "Do it as quickly as possible".

In this chapter, four major points that are concerned with the aim and findings of this project will be discussed: (a) methodological sophistication, (b) individual differences, (c) task completion time and (d) ultimate attainment of native-like lexical knowledge and organisation.

First, an assessment of the replication of Haastrup and Henriksen's (2000) sorting tasks and the sorting tasks that we developed in this project will be made. This includes addressing the question of whether the participants sorted the words into groups according to related meanings irrespective of the word types they were given. For this purpose, we will re-examine the mixed word sorting task results, which were reported in Chapter 4, from this perspective. Second, individual differences in aspects of the sorting task results (e.g., task completion time, cluster number, size and variability and individual and group dendrograms) will be addressed. The results across sorting tasks showed that individual differences as observed in the bulk of measured *SD*s were always fairly large. We will discuss this issue in view of the variability and consistency the results produced, individual level analyses and word type effects and the relationships between individual data analyses and group data analyses. Third, we will discuss the task completion times participants took to complete the sorting tasks. The discussion will include within-subject comparisons of task completion time across tasks given under the same design except in tested words. Then we will address why L2 participants produced the biggest task completion time variability in completing the 1K *Treasure Island* verb sorting task, why both L1 and L2 participants took the longest time to complete the 1K *Treasure Island* adjective sorting task, and what is important about the fact that difference of task completion time between the two groups was the smallest in the 1K *Treasure Island* noun sorting task. As the fourth and final discussion point, the ultimate attainment of native-like lexical knowledge and organisation will be addressed. We will particularly address the findings we made in Chapter 8 that 1K *Treasure Island* nouns produced statistically significant L1 and L2 differences in the experiment and that English loanwords played a role in the results. These findings suggest that even for advanced level speakers of English as an L2 there remain facets of lexical knowledge and organisation that are difficult to master. We will discuss the plausible factors that might account for the difficulties in attaining native-like lexical knowledge and organisation.

9.2 Methodological sophistication

In this section, three issues will be addressed: (a) Haastrup and Henriksen's (2000) and the present project's sorting tasks, (b) validating sorting tasks as techniques for lexical organisation research and (c) the effects of using a cohesive *Treasure Island* passage in selecting tested words and the sorting task direction of "Do it as quickly as possible".

9.2.1 Haastrup and Henriksen's (2000) and present project's sorting tasks

The replication study of Haastrup and Henriksen (2000), which was reported in Chapter 3, revealed that their sorting tasks (i.e., the Primary Sorting Task and the Card-Sorting Task) had more to do with how well participants knew word meanings rather than with how participants developed lexical organisation. This led to the weakened validity of their sorting tasks as data elicitation techniques for lexical organisation research.

In the Primary Sorting Task, participants were asked to place 39 adjectives of emotion and physical dimension into one of four categories: HAPPY, AFRAID, WEIGHT and SIZE, or TEMPERATURE. In the Card-Sorting Task, they were directed to sort 30 adjectives into four categories of adjectives of emotion (AFRAID, ANGRY, HAPPY or SAD). A serious limitation of these tasks lay in the fact that the tasks included words that were low-frequency, rare words (e.g., *chuffed, elated, distressed* and *grumpy*). Both "near-beginner" Danish school children, participants in Haastrup and Henriksen's study, and the intermediate-level Japanese college EFL learners in the replication study we conducted certainly had little knowledge of these words and were unlikely to have encountered them before. This fact was reflected in the results, which showed that both groups failed to sort them into the correct categories and achieved low scores, compared with L1 controls. The sorting tasks Haastrup and Henriksen developed were not reliable enough to examine the non-native speakers' lexical organisation. One of the requisites in developing a psycholinguistic data elicitation technique in lexical organisation research is to select words for the task that participants know the meanings of well. Without this, the task would tap into something else and fail to accomplish its mission. Accordingly, for the sorting tasks developed in this thesis project, we used high frequency words, those being the first 500 or 1000 words taken from the JACET 8000 Word List. These are words that advanced-level Japanese EFL learners should have no trouble in processing their meanings and sorting them into groups semantically. The results reported in Chapters 4 through 8 showed that using high frequency words worked well to achieve the aim of each experiment.

Another problem that Haastrup and Henriksen's sorting tasks had was concerned with the task directions. Particularly in the Card-Sorting Task participants were told that the tested words were all adjectives of emotion. Moreover, participants were directed to sort them into four groups of adjectives of emotion (i.e., SAD, HAPPY, ANGRY or AFRAID) and to leave unsorted adjectives by themselves. With these directions for the task, participants got hints that helped them to carry out the task. This excessive help degraded the task's reliability as one for lexical organisation research. Moreover, fixing in advance the number of sorting categories participants were going to make made it impossible to address the issue of whether there were individual differences in the cluster number and size participants produced. It has been shown that fixed sorting tasks have "the advantage of standardizing the variance of the sorting categories" (Coxon, 1999, p. 20). However, in examining the cluster structure of L1 and L2 lexical organisations, fixing the category number in sorting has more disadvantages than advantages. Eventually, such tasks lack the sensitivity to examine how large individual differences in cluster number and size would be among participants and whether the differences would be substantially large between the NS and NNS groups. Thus, in the sorting tasks we developed, participants were directed to sort words freely (i.e., there was no limitation imposed in the number of sorted words they were making). With this revision, the sorting tasks had the validity to address the cluster number, size, variability and organisation, which have been the central areas of research throughout this thesis project.

After examining Haastrup and Henriksen's study and the replication we conducted in the early stages of this project, we determined not to use the results of native-speakers' sorting task as baseline data in deciding scoring keys for non-native speaker performance. This decision was made for two reasons. First, the results of native-speakers were not consistent enough to be reliable baseline data as Haastrup and Henriksen themselves reported. It should be noted that this problem has been noted by other researchers of L2 lexical organisation as well (see, for example, Fitzpatrick, 2007, 2008; Meara & Schur 2002; Wolter, 2005). Second, and more importantly, what is worthwhile to address regarding L1 and L2 lexical organisations is to compare the NS sorting task results to their NNS counterpart's and to examine the differences between them. Therefore, in a series of sorting tasks after the replication, we did not analyse the NNS results as "correct or wrong" answers by using native-speaker performance criterion. Instead we compared the results of two participant groups to each other and searched for differences in certain aspects of the data. After comprehensively reviewing the experiments reported in the previous chapters, this decision seemed to have worked well in that the sorting tasks succeeded in revealing crucial differences in cluster structures of L1 and L2 lexical organisations.

9.2.2 Validating sorting tasks as lexical organisation research techniques: Re-examination of 0.5K mixed word sorting task results

When the sorting tasks were given to NS and NNS groups using 1K *Treasure Island* verbs, adjectives and nouns, participants sorted words into groups according to their meanings and the results had the sensitivity to distinguish native speakers from non-native speakers regarding lexical organisation (see Chapters 6, 7 and 8). Particularly, the final clusters identified in the group dendrograms revealed that the NS group always produced a larger number of clusters with their sizes being smaller than those of the NNS group. This was interpreted to mean that L1 participants were more aware of lexical items as distinct sets than their L2 counterparts. In addition, the composing lexical items in respective clusters were evidently different between the two groups. Thus the L1 and L2 lexical organisations were structurally different from each other. Sorting tasks using 1K *Treasure Island* verbs, adjectives and nouns proved to be valid techniques in probing into L1 and L2 differences in lexical organisation. The remaining crucial issue is whether the sorting tasks in general worked well as predicted, irrespective of the word type tested. In this regard, we have left the 0.5K mixed word sorting task results unanalysed, although the descriptive statistics-based results were reported in Chapter 4. That is, we have not addressed the question yet: "Did participants still sort the mixed words into groups according to related meanings?" To answer this, it is necessary to confirm whether a sorting task using 0.5K words taken from a mixture of word classes produced results in line with those of 1K *Treasure Island* verbs, adjectives and nouns. It seems essential for us to address this task validity issue at this final stage of the project when all the planned experiments have been completed and we have made the findings of L1 and L2 differences in lexical organisation that should lead to possible further studies using sorting tasks.

Keeping this purpose in mind, a cluster analysis was run on the 0.5K mixed word sorting task results and the final clusters were identified for the NS and NNS groups. The analysis followed exactly the same procedure we used for the 1K *Treasure Island* verbs, adjectives and noun sorting tasks. The NS results are shown in Table 9.1 and the NNS results in Table 9.2.

Table 9.1. Mixed word sorting: Final clusters identified for the NS group ($k = 50$)

No.	Cluster	Word No.	Words
1	Dimension of quantity and time	10	*nothing, very, all, also, early, next, while, then, already, ago*
2	Dimension of power and history	10	*law, president, police, doctor, country, war, power, history, century, arm*
3	Dimension of state and place	8	*open, close, lot, business, shop, street, place, area*
4	Dimension of nature	5	*matter, form, clear, nature, air*
5	Dimension of human	4	*our, figure, person, boy*
6	Dimension of movement and cognition	13	*walk, step, arrive, keep, find, social, help, cry, stand, dream, reason, understand, believe*

Table 9.2. Mixed word sorting: Final clusters identified for the NNS group ($k = 50$)

No.	Cluster	Word No.	Words
1	Dimension of quantity	5	*also, very, lot, nothing, all*
2	Dimension of history and time	8	*history, century, while, early, already, ago, then, next*
3	Dimension of state, movement, place and nature	13	*open, close, stand, walk, arrive, step, shop, street, place, area, clear, nature, air*
4	Dimension of power, cognition and human	24	*our, country, social, war, power, law, business, president, police, doctor, person, boy, help, cry, keep, understand, find, dream, believe, reason, matter, form, figure, arm*

Table 9.1 shows that the final clusters the NS group produced were less distinctive than the ones produced for the 1K *Treasure Island* verb, adjective and noun sorting tasks. For example, the participants merged the dimension of quantity (*nothing, very, all, also*) with the dimension of time (*early, next, while, then, already, ago*) into a final cluster. Broadly speaking, these two dimensions share some semantic field (e.g., abstract concept of measurement) with each other and form a single dimension. Nonetheless, the merged cluster was not a clear-cut one for a single cluster structure in the mental lexicon. The lack of this distinctiveness appears to be caused by the fact that the tested words were very high frequency 0.5K words which were selected randomly from the JACET 8000 Word List. Accordingly, even if they were content words, they (e.g., *next, then, already*) seemed to be integrated in participants' lexical organisation as lexical items which carry less distinctive semantic meaning than others. This also held true with the pronoun *nothing* and the conjunction *while*, which were also components of the merged dimension of quantity and time cluster. Tables 9.1 and 9.2 showed

that this same problem of having a lack of distinctiveness was also found in the dimension of power and history cluster, the dimension of state and place cluster and the dimension of movement and cognition cluster.

Meanwhile, the results provided firm evidence showing that the sorting tasks themselves were valid data elicitation techniques to probe into lexical organisation. This was observed in the distinctiveness of the dimension of nature cluster (*matter, form, clear, nature, air*) and the dimension of human cluster (*our, figure, person, boy*). It should be noted that the components of these two final clusters were comprised of lexical items of mixed word classes, but they still formed a single cluster with a shared semantic field such as *nature* or *human*.

Given these results together, it seems reasonable to assume that participants sorted the tested words into groups according to their meanings and irrespective of their word types. However, very high frequency words and function words had little ability to produce meaningful results, and they were often difficult to interpret as shown above. A comparison of the NS results (Table 9.1) and NNS results (Table 9.2) revealed another promising merit of the sorting task in that NS produced more final clusters ($n = 6$) than NNS did ($n = 4$). This is consistent with the results of the 1K *Treasure Island* verb, adjective and noun sorting tasks reported in Chapters 6, 7 and 8, where the sorting tasks always produced the result that NS produced a larger number of clusters than NNS did. Thus, we may conclude that the sorting tasks were reliable enough to distinguish L2 participants from their L1 counterparts in that native speakers were more aware of lexical items as being distinctive sets than non-native speakers. Finally, it is cautioned that future research using sorting tasks should not contain such words as very high frequency words, function words and content words having weak semantic distinctiveness, which would not produce reliable results in probing into L1 and L2 lexical organisations by means of the semantic clustering behaviour of the participants.

9.2.3 Task sensitivity: Revisions of word selection procedure and task direction

In this section, we will examine whether the revisions we made on the sorting tasks had an effect on the sensitivity of the method. For this purpose, all five sorting tasks we developed will be examined while comparing their power to distinguish L1 participants from L2 participants in (a) distances of individual participants' dendrograms, (b) variability of individual participants' dendrograms and (c) distances of NS and NNS group dendrograms. The revisions we are examining are concerned with the word selection procedure by introducing a cohesive *Treasure Island* passage and directing participants to complete the task as quickly as possible. These revisions were made on the sorting task reported in Chapter 6 (1K verb sorting) and the adjective and noun sorting tasks reported in Chapters 7 and 8 that followed the same approach as the 1K verb sorting task. Therefore, the point of the examination in this section lies in comparing the aspects of the results before and after the revisions.

In Chapter 6, we found that by changing the task direction to "Please do it as quickly as you can", L2 participants were not statistically significantly different from their L1 counterparts in task completion time although they still took on average a slightly longer time in task completion. Thus the reliability of the sorting task was boosted in that both groups activated and processed lexical knowledge and organisation in a similar amount of time. Comparative analysis of task completion time between 1K *Treasure Island* verb, adjective and noun sorting tasks in Chapter 8 made some interesting findings (e.g., both NS and NNS groups took longest in adjective sorting, which will be discussed below in 9.4). We will now start discussing (a) the effects of the task revisions by examining the power to tell L1 participants from L2 participants in individual participants' dendrograms. Table 9.3 shows the means and *SD*s of the distances of individual participant's dendrograms (distance matrices) that the sorting tasks produced.

Table 9.3. Distance of each individual participant's dendrogram

Sorting task	NS		NNS		*t*-value	*F*-value
	Mean	*SD*	Mean	*SD*		
0.5K JACET	8.11	2.95	7.54	3.50	0.66n.s.	1.41n.s.
1K JACET verbs	8.93	3.05	8.33	3.41	0.72n.s.	1.25n.s.
Treasure verbs	7.40	2.87	7.70	2.91	0.40n.s.	1.03n.s.
Treasure adjectives	7.73	2.65	7.27	2.27	0.73n.s.	0.74n.s.
Treasure nouns	7.93	2.23	6.47	2.06	2.65**	0.86n.s.

Note. Both NS and NNS groups had 28 participants in the 0.5K JACET sorting task and 30 participants for the other sorting tasks. 0.5K JACET = The sorting task using words selected randomly from the most high frequent 500 words in the JACET 8000 Word List; 1K JACET verbs = The sorting task using verbs selected randomly from the most high frequent 1000 words in the JACET 8000 Word List; *Treasure* verbs = The sorting task using 1Kverbs in the JACET 8000 Word List taken from a cohesive *Treasure Island* passage; *Treasure* adjectives = The sorting task using 1K adjectives in the JACET 8000 Word List taken from a cohesive *Treasure Island* passage; *Treasure* nouns = The sorting task using 1K nouns in the JACET 8000 Word List taken from a cohesive *Treasure Island* passage. The same abbreviations are used in subsequent tables in this chapter. $^{**}p < 0.01$; n.s. = not significant.

Table 9.3 shows that the revisions made on the 1K JACET verb sorting task, which were in effect for the sorting tasks using *Treasure Island* verbs, adjectives and nouns, brought about distinctive L1 and L2 differences solely in the *Treasure Island* noun sorting task. That is, L2 participants made a statistically significant smaller number of clusters only in the noun sorting task. There was no such difference in the verb sorting and adjective sorting results, suggesting that in these two tasks NNS were not substantially different from NS in the individual dendrograms they made. This was consistent with the findings we made on the L1 and L2 differences in cluster number, size, number of single, isolated words and mean largest cluster. Regarding these variables as well, the two groups were substantially different from each other only in the results of 1K *Treasure Island* noun sorting. Examining these results together, it can be reasonably concluded that the revisions conducted on the sorting task on the 1K JACET verb sorting did not produce a direct impact on all of the subsequent experiments. However, they made the tasks robust and boosted task reliability in examining underlying cluster structures of the L1 and L2 mental lexicons in that both participants groups completed the task within an almost similar time duration. Eventually, we discovered that nouns can be the predictors of the L1 and L2 differences in lexical organisation in the final experiment.

To address the issue of (b) variability of individual participants' dendrograms, Table 9.4 shows the average dispersions of individual participants' dendrograms against the dendrograms of the group data (where the smaller the value is, the better the individual results fit with the group data) and the *p*-values regarding the L1 and L2 differences.

Table 9.4. Significance levels (*P*-values) of permutation test results: Average dispersions of dendrograms of individual participants against the dendrogram of the group data

Sorting task	Average dispersion		*p*-value
	NS	NNS	
0.5K JACET	109.41	113.11	0.413
1K JACET verbs	110.68	104.28	0.653
Treasure verbs	114.65	106.36	0.666
Treasure adjectives	103.66	106.21	0.413
Treasure nouns	71.48	95.92	0.017*

Note. Average dispersion = Average dispersion of individual participants' dendrograms against the dendrogram of the group data; *p*-value = significant level of the difference of the average dispersion between the NS and NNS results as calculated by a permutation test which was administered on the assumption that the average dispersion of the NS group's individual dendrograms and the NNS's would be equal to each other; *$p < 0.05$.

Table 9.4 shows that the revisions made to the 1K JACET sorting task did not bring about an immediate impact on the *Treasure Island* verb sorting task in that the new sorting task failed to produce a statistically significant L1 and L2 difference ($p = 0.666$). This was also the case with the sorting task using 1K *Treasure Island* adjectives ($p = 0.413$). However, the sorting task using 1K *Treasure Island* nouns produced a significant L1 and L2 difference, where the NS group's average dispersion of individual participants' dendrograms was distinctively smaller than its NNS counterpart's ($p = 0.017$). This result is completely consistent with the results of the distances of individual participants' dendrograms examined above in that although the task revisions did not produce a direct impact on all the following tasks, it contributed to reveal the L1 and L2 differences in lexical organisation which was firmly detected using 1K *Treasure Island* nouns. This is further evidence for the positive effects of the revisions made on the 1K JACET verb sorting task.

Addressing whether the revisions made on the sorting task increased the sensitivity to distinguish (c) the NS group dendrogram from its NNS counterpart's, relevant data was tabulated and is shown in Table 9.5.

Table 9.5. Significance levels (P-values) of the NS and NNS group dendrogram differences

Sorting task	Significance level (Permutation test result on the group dendrograms)
0.5K JACET	0.0176
1K JACET verbs	0.4960
Treasure verbs	0.0120
Treasure adjectives	0.0012
Treasure nouns	0.0388

Table 9.5 reveals that the revisions brought about an immediate, distinct improvement on the sorting task. The results of the 1K JACET verb sorting task, the pre-revision sorting task which used 1K JACET verbs, did not produce a statistically significant difference between the NS and NNS group dendrograms ($p = 0.4960$). Meanwhile, all the sorting tasks after the revisions were made generated significant differences as shown in the result of sorting tasks using *Treasure* verbs ($p = 0.0120$), *Treasure* adjectives ($p = 0.0012$) and *Treasure* nouns ($p = 0.0388$). It should be noted that the sorting task using 0.5K JACET words also brought about a significant L1 and L2 difference ($p = 0.0176$). This suggests that at the group level the 0.5K JACET word sorting task had the validity to probe into the lexical organisation of L1 and L2 mental lexicons. Even if the tested words were comprised of mixed word classes, participants still produced sorting task results where L1 participants were distinctively different from their L2 counterparts (see the previous section examining the 0.5K mixed word sorting task results). However, this was not the case with the 1K JACET verb sorting task. Some intra-linguistic features of sorting English verbs seemed to play a role, which will be addressed in section 9.4.

Examining these results as a whole, we may conclude that the revisions made on the sorting tasks functioned satisfactorily to meet the present research purpose of probing into the differences of L1 and L2 lexical organisations. The revised sorting tasks were reliable in tapping into these organisations and succeeded in finding the predictors of the L1 and L2 differences. In addition, it should be noted that the adoption of multivariate analyses, particularly cluster analysis, seemed to reveal the underlying structural differences more reliably. By examining the number and components of final clusters as a result of cluster analysis, we found that L1 and L2 lexical organisations were actually different from each other in important respects. This would not have been possible by merely relying on classical descriptive statistics-based analyses, which includes analysis using a cut-off point where we searched for "native-like" linked pairs in the NNS results (Chapter 5). Theoretically, with 50 words used in a sorting task, there could have been 1,225 potential pairs of linked words in the results. However, the cut-off point approach detected only 41 "native-like" linked pairs, which only explained an extremely small portion of the total organisation. It should also be noted that permutation tests enabled us to shed fresh light on the cluster analysis data, with which we could reliably test whether there was a distinctive difference in individual participants' and group dendrograms and the variability between the two groups. This also seems to have helped us to advance this project as being applicable to future lexical organisation research.

9.3 Individual differences

In this section, issues of individual differences in the sorting task experiments will be addressed. We will particularly discuss (a) between-subject variability and within-subject consistency, (b) individual level analyses and word type effects and (c) the relationship between individual data analyses and group data analyses.

9.3.1 Individual differences in sorting task results: Variability and consistency

The sorting task results reported and examined in this thesis project produced rather large variability throughout all the analyses. Descriptive statistics-based analyses (e.g., cluster number, size, number of single, isolated words, and mean largest cluster) and multivariate analysis- and permutation test-based analyses (e.g., distances of each individual participant's dendrogram, average dispersions of dendrograms of individual participants against the dendrogram of the group data and distances between the NS and NNS group dendrograms) revealed that the values were impressively large across the results that both L1 and L2 participant groups produced. For example, the average distance of the dendrograms L1 participants produced for the 1K JACET verb sorting task was 8.93 and the *SD* was 3.05, whereas the L2 average distance was 8.33 and the *SD* was 3.41 (see Table 9.3). This seems to suggest that a sorting task is a type of cognitively demanding 'difficult' data elicitation technique where "individual differences appeared to become more prevalent as the difficulty of problems increased" (Chronicle, MacGregor, Lee, Ormerod & Hughes, 2008, p. 41) in nature. The point is, however, that there were no statistically significant differences in variability between the NS and NNS groups (see *F*-values of the sorting task results tabulated in Table 9.3). This task stability allowed us to conduct analyses in a reliable way, which led to the findings of L1 and L2 lexical organisations reported in Chapters 3 to 8.

The most crucial issue of individual differences in psycholinguistic experiments lies in whether individual participants who produced large variability between tasks still consistently carried them out. That is, did individual participants complete the sorting tasks in a stable and consistent manner in spite of the between-subject variability across the tasks? If they did, the tasks can be regarded as reliable ones. In this regard, we need to examine whether the sorting tasks we gave in this project were stable enough as a data elicitation tool. In answering this issue, Fitzpatrick (2007) offers us a helpful hint. Addressing whether adult native English speakers can be considered reliable in terms of 'native-like' responses in word association tests, she administered matching word association tests a week apart to a group of the same adult native English speakers. The results showed that participants made highly varied responses as a group but that individual participants responded in a remarkably consistent way. This suggests that there was between-subject variability among the participants but that there was also reliable within-subject consistency in the word association behaviour of adult English native speakers. Thus, the question worth addressing for us now is whether the sorting tasks in this project (which produced a large amount of between-subject variability as was the case with Fitzpatrick's WAT-based research) produced reliable within-subject consistency as well.

Limitations of the within-subject consistency examination of the present sorting tasks should be addressed before reporting the results. In this project it took us five years to carry out the six experiments, those being the replication of Haastrup and Henriksen (2000) and the five sorting tasks that were developed and revised. This was attributed to the fact that I myself gave each task to individual participants (30 NSs and 30 NNSs = 60 for each experiment, but 28 NSs and NNSs for the 0.5K mixed word sorting task) in each experiment to be able to conduct the tasks in a consistent fashion. Thus it took me about three months to complete each experiment. It should be noted that because of this limitation the examination we are making in this section is not as robust as Fitzpatrick (2007). As reported above, major revisions were made to the sorting task on the 1K JACET verb sorting task, the task which was done in the third year, and the three sorting tasks after it were done under the same design except in the type of tested words. Accordingly, we will address the issue of this within-subject consistency while examining the last three sorting tasks.

Among all the participants, there were 11 L1 participants and 10 L2 participants who completed all the three sorting tasks (i.e., the 1K *Treasure Island* verb, adjective and noun sorting tasks). To examine whether within-subject consistency was maintained across the three sorting tasks, I ran a correlation analysis on the results of the number of clusters (which excluded single, isolated words) that the NS and NNS groups made. Table 9.6 shows within-subject correlations among the three sorting tasks for the NS group.

Table 9.6. NS's within-subject correlations among the sorting tasks ($n = 11$)

	Correlation coefficient	p-value
Verbs vs. adjectives	-0.236	0.495
Verbs vs. nouns	-0.010	0.978
Adjectives vs. nouns	0.937	< 0.001

Table 9.6 shows that there was an extremely high correlation between the adjective and noun sorting tasks ($r = 0.937$, $p < 0.001$) for L1 participants. There was no correlation between the verb and adjective sorting tasks nor the verb and noun sorting tasks. These results revealed that native speakers of English produced consistent sorting task results as long as the tested words were adjectives or nouns. However, this was not the case with the verb sorting task, where the results had little predictability of the adjective and noun sorting task results.

Table 9.7 shows within-subject correlations among the three sorting tasks for the NNS group.

Table 9.7. NNS's within-subject correlations among the sorting tasks ($n = 10$)

	Correlation coefficient	p-value
Verbs vs. adjectives	-0.054	0.887
Verbs vs. nouns	-0.459	0.189
Adjectives vs. nouns	0.577	0.082

Table 9.7 reveals that there was a moderate correlation between the adjective and noun sorting tasks ($r = 0.577$) for L2 participants. It is noted that the NNS group failed to reach the 5% significance level ($p = 0.082$) by a small margin. These results suggest that L2 participants were less consistent (even in the sorting tasks they maintained consistency) than their L1 participants were. In the cases of other pairs of sorting tasks, i.e., verb vs. adjective sorting and verb vs. noun sorting, L2 participants failed to maintain within-subject consistency, as judged from the correlation coefficients and p-values. The verb sorting results had little ability to predict the results for L2 participants of other sorting task.

Analysing these results as a whole, although the examination was limited, it is highly plausible that both L1 and L2 participants behaved consistently in the adjective and noun sorting tasks. Moreover, L1 participants maintained much higher consistency across the sorting tasks than L2 participants did. Native speakers were stable and predictable in their sorting behaviour of adjectives and nouns. Meanwhile, verbs produced a large variability in the sorting task among both the NS and NNS groups. As a result, it seems to be difficult to maintain within-subject consistency with verbs.

Needless to say, the analyses of within-subject consistency and variability we did in this section have no effect on the analyses conducted in the previous chapters. The sorting tasks always produced large between-subject variability, but the variability was always non-significant and the subsequent analyses were done in a reliable way. The primary purpose of this project is to examine L1 and L2 differences of cluster structure in lexical organisation using sorting tasks. Thus principal analyses of the results have been done by means of between-group (subjects) comparisons. All the analyses we made had no problem in this regard. It should be noted that even the results of 1K *Treasure Island* verbs, which produced the largest within-subject variability in the NS and NNS results, were reliably analysed and revealed aspects of L1 and L2 differences by means of descriptive statistics-based analysis, cluster analysis and permutation tests.

9.3.2 Individual level analyses and word type effects

In this section, we will address the issue of individual level analyses by means of their relationships to word type effects on sorting tasks. This project revealed that only 1K *Treasure Island* nouns were found to be predictors of L1 and L2 lexical organisations when individual level analyses were made. Table 9.8 summarises the main variables which turned out to be statistically significant in the sorting task results, which were concerned with individual level analyses and the differences between L1 and L2 group dendrogram distances.

Table 9.8. Variables that produced distinctive differences between L1 and L2 lexical organisations

	Cluster number	Cluster size	Largest cluster	Individual dendrogram	Group dendrogram
0.5K JACET					√
1K JACET verbs					
Treasure verbs					√
Treasure adjectives					√
Treasure nouns	√	√	√	√	√

Note. Cluster number = mean cluster number (which excludes single, isolated words); Cluster size = mean number of words per cluster (which excludes single, isolated words); Largest cluster = mean largest cluster participants made; Individual dendrogram = mean distance of each individual participant's dendrogram; Group dendrogram = difference between L1 and L2 group dendrogram distances. The symbol √ means that the variable revealed a statistically significant L1 and L2 difference.

A glance at Table 9.8 shows that L1 and L2 differences in aspects of lexical organisation were not readily detected in most experiments. By the time the *Treasure Island* noun sorting task was carried out, we had detected no statistically significant L1 and L2 difference in any of the variables we had been addressing except in the group dendrogram. Of course, our sorting tasks almost always produced results which suggested that L1 participants tended to make a larger number of clusters and fewer words per cluster. However, their differences always failed to reach a statistically significant level except in the 1K *Treasure Island* noun sorting task. Thus, L2 participants appear to have integrated the basic lexical knowledge of 1K words into their L2 lexical organisation to a high degree at least regarding the aspects which the present sorting tasks tapped into. Meanwhile, cluster analyses showed that even though L1 and L2 participants were not different from each other in cluster number and size (except in 1K *Treasure Island* noun sorting), their lexical organisations were substantially different from each other (including 1K *Treasure Island* noun sorting). This suggests that it might not be so difficult for non-native speakers of English (i.e., advanced-level Japanese learners of English) to learn and assimilate the meanings of high frequency English words. But this is not the case with the attainment of native-like lexical organisation. Our results suggest that it is very difficult for L2 learners to master native-like lexical organisation. This mismatch of the attainment in L2 learners' lexical knowledge and organisation was most evidently shown by the 1K *Treasure Island* nouns, which can be reliable predictors of distinct L1 and L2 differences (including those in individual dendrograms).

9.3.3 Individual data analysis vs. group data analysis

The last column of Table 9.8 shows that L1 and L2 group dendrograms were found to be distinctly different from each other (except in the 1K JACET verb sorting task) when the data was submitted to permutation tests. At first glance, this appears to be somehow contradictory to the results of individual dendrogram analyses (shown in the second to the last column) where there was no statistically significant difference between the two groups (except in 1K *Treasure Island* noun sorting). As a last issue to be addressed regarding individual differences in the sorting task results in this section, the relationship between individual data analysis and group data analysis will be discussed. We will also discuss the adoption of computer modelling in future studies as a promising approach to cope with individual difference problems in psycholinguistic experiments.

Distances of inter-object relationships in an individual participant's dendrogram (i.e., sorting task results) are represented by binary values of "zero" or "one". A value of "zero" means a perfect similarity between the lexical items in question and a value of "one" shows a perfect dissimilarity between them. In Chapter 7 discussing 1K *Treasure* adjective sorting results, we indicated that in individual dendrogram analysis, this binary-based data structure of the individual sorting task results makes it extremely difficult to detect individual differences when individual results were examined solely by themselves. Underlying structures of this type of data are not easily detected by the analysis, even if there are some meaningful features. Table 9.8 shows that this held true with other sorting task results as well and the differences in individual L1 and L2 dendrograms did not reach a statistically significant level. Meanwhile, when these individual results were aggregated and submitted to multivariate analysis (i.e., cluster analysis) and the following permutation test to examine the differences of L1 and L2 group dendrograms, underlying differences were clearly unearthed. This is the merit of group data analysis in examining sorting task results.

As explained above, latent individual differences can be revealed by group data analysis, which has been adopted widely by psycholinguistic data analysis. Miller (1969), Preece (1976), Rapoport and Fillenbaum (1972) and Routh (1994), all of which we reviewed in Chapter 2, were in line with this approach and carried out multivariate analyses of the psycholinguistic data they collected. Another reasonable explanation is that meaning and its structure in memory is not confined to individuals, but rather is a property of a culture (Deese, 1965; Preece, 1976). By aggregating and examining individual participants' data, we can gain an invaluable insight into its meaning and structure, which is perhaps not the simple sum or average of individual results in that multivariate analysis allows us to examine the underlying organisation. Preece states that "pooling individual data will produce more valid results" (p.5). Our group data analysis by means of cluster analysis and permutation tests to examine significant L1 and L2 differences was carried out while keeping this understanding of individual data in mind.

Although it is beyond the scope of the present project, computer modelling (i.e., simulation) is a promising approach to lexical organisation studies. One of the merits of computer modelling is that the approach can reliably test assumptions which are free of individual difference parameters by postulating a model of, for example, optimal native speaker lexical organisation and examining validity through simulations. As we have seen, human cognitive structure is not simple and individual differences are always found in participants' behaviour in data elicitation tasks. By building such a model of a lexical network, we can test its validity again and again while changing the parameters where needed. That is, computer simulations can postulate this optimal lexical organisation that is free of real-world human individual difference variables (e.g., personality, language aptitude, motivation, leaner strategies and learning style) (see, for example, Dörnyei, 2006; Dörnyei & Skehan, 2003; Singleton, 1999; Skehan, 1991, for studies of individual differences and second language acquisition). Simulations can be done virtually as many times as needed until the results meet the theoretical assumptions we make. In relation to the findings of the present project, for simulations it is possible to build optimal models of pseudo-L1 and L2 lexical organisations which each have a small lexicon of 50 lexical items. The model entails "only the essential features of a network" (Meara, 2006, p. 638). The model's plausible sorting behaviour can be tested by comparing it to the results of the sorting tasks we collected from human participants. Through the interactions between the computer modelling results and the human participants' data, we may deepen our understanding of L1 and L2 differences in lexical organisation. Larsen-Freeman and Cameron (2008) stated that computer simulations or models offer "an important approach to researching complex dynamic systems. Although still in its infancy, modeling in applied linguistics shows great promise" (p. 247). Meara (2004b, 2006, 2007a, 2007b), Meara and Schur (2002), Schur (2003, 2007), Wilks (1999), Wilks and Meara (2002), Wilks, Meara and Wolter (2005), some of which we reviewed in Chapter 2, are pioneering studies of modelling. These studies have made important findings of mental lexicons using types of word association tests and computer simulations. It is expected that sorting task experiments will also benefit from computer simulations and the validity of the findings that the present project have made will be tested through simulations free of individual difference variables.

9.4 Task completion time

In this section, two questions relating to task completion time will be addressed. First, we will answer the question of which variables contributed to producing significant within-subject differences in task completion time. Second, we will answer the question of what factors led participants to generate some of the characteristics regarding task completion time. We will examine the 1K *Treasure Island* verb, adjective and noun sorting tasks one by one.

9.4.1 Task completion time and sorting task results: Within-subject comparison

In Chapter 8, making within-subject comparisons of task completion time between 1K *Treasure Island* verb, adjective and noun sorting tasks, we found that both NS and NNS groups took the longest time in completing the adjective sorting task. It was also found that the difference in task completion time between the two groups was the least in the noun sorting task, and among L2 participants the difference was the largest in the verb sorting task. In this section, we are particularly interested in the fact that the two groups took the longest in the adjective sorting task, which suggests that some relevant sorting behaviour of the L1 and L2 participants might have played a role. Thus our primary question is whether both groups exhibited some similar sorting behaviour in the completion of the adjective sorting task. It is assumed that evidence for L1 and L2 similar sorting behaviour can be revealed by examining the sorting task variables of mean cluster number, words per cluster and largest cluster. It is predicted, for example, that both NS and NNS groups might have made the mean largest cluster number or alternatively the mean smallest cluster number among the three sorting tasks as a reflection of their slowest processing time in sorting adjectives. There is no theoretical basis to hypothesise which of these theories would be plausible before fully analysing the data. However, it is highly possible that both groups would have behaved in a similar fashion and produced the same results. If we could find some result showing L1 and L2 similarity on this issue, that would be evidence for the plausibility that L2 lexical organisation is similarly activated and processed in a specific type of word class (i.e., 1K adjectives) as is L1 lexical organisation. On the contrary, if we fail to find it, that would suggest that some other factor might have played a role in L1 and L2 similarities in the task completion time for adjective sorting.

Accordingly, three sets of data, whose between-subject (group) analyses were made in Chapters 6, 7 and 8 respectively, were examined while addressing whether within-subject differences would be found between the results of 1K *Treasure Island* verb, adjective and noun sorting tasks. As stated above, the variables examined were mean cluster number, mean number of words per cluster and mean largest cluster.

Repeated-measures ANOVAs run on the data (details omitted) found no evidence in any of the three variables (i.e., mean cluster number, mean words per cluster and mean largest cluster) that the NS and NNS groups behaved similarly in task completion time in the adjective sorting task. These three variables did not seem to be indicators which reflected that both groups took the longest time in adjective sorting. Our analyses also failed to find any evidence related to the result that the difference of task completion time between the two groups was the least in the noun sorting task, and among L2 participants the difference was the largest in the verb sorting task. Thus, the tendencies that we detected regarding task completion time cannot be attributed to the variables of cluster number, words per cluster and largest cluster participants produced. Some other factors (i.e., intra-lexical features of 1K verbs, adjectives and nouns that were used in the experiments) appear to have played a role in producing the results concerning task completion time. This issue will be discussed in the next section.

9.4.2 Intra-lexical features affecting task completion time

In this section, we will attempt to answer what intra-lexical features of the word types in the sorting tasks might have played a role in producing each of the results in task completion time characteristic to the 1K *Treasure Island* verb, adjective and noun sorting tasks. We will specifically address three questions: (a) Why did L2 participants produce the biggest task completion time variability in completing the 1K *Treasure Island* verb sorting task? (b) Why did both L1 and L2 participants take the longest time to complete the 1K *Treasure Island* adjective sorting task? and (c) What is important about the difference of task completion time between the two groups being the smallest in the noun sorting task?

(a) *Why did L2 participants produce the biggest task completion time variability in completing the 1K Treasure Island verb sorting task?*

To answer this question, we should pay particular attention to the participants' comment in sorting 1K verbs that they often found it difficult to decide which meaning of a verb they should decide on (see 5.4.1 Intrinsic complexities of verbs and the revised sorting task). The high frequency verbs used in the 1K *Treasure Island* verb sorting (e.g., *find, take, leave, show, face, strike, teach*) were polysemous and thus have different meanings according to what types of words follow or come before them (e.g., concrete or abstract nouns, animate or inanimate nouns, human-beings or non-human-beings). This nature of high frequency verbs has been noted by Miller and Fellbaum (1991): "Verbs are more polysemous than nouns: the nouns in *Collins* [*English Dictionary*] have on the average 1.74 senses, whereas verbs average 2.11" (p. 214) (see also Gentner, 1981, for the same conclusion). Thus, participants likely activated more than one meaning of each of the high frequency verbs tested in the sorting task experiment and the meaning they each decided on while completing the task must have varied from participant to participant to a significant degree. Regarding task completion time, this variability also played a role, where some participants took a shorter time than others and some took longer than others, with participants revealing large individual differences. Eventually, they produced an impressively large variability in task completion time, and the magnitude was bigger than when they worked with the adjective and noun sorting tasks.

As Källkvist (1998) states, English verbs have "more complex semantics and vary more cross-linguistically than nouns" (p. 150). We should be prudent in addressing the factors of L2 participants' variability in task completion time. However, given the polysemous nature and semantic complexity of English verbs, it seems highly possible that high frequency English verbs are the word type that is most difficult for non-native speakers to master (see also Källkvist, 1999; Lennon, 1996; Miller & Fellbaum, 1991; Read, 2004). When L2 participants work on a psycholinguistic experiment such as a sorting task, dissimilarities manifest regarding their mastery of various aspects of verbs and individual differences in how long it takes them to decide on the meaning of words. Eventually, L2 learners appeared to vary significantly in task completion time as the results showed.

(b) *Why did both L1 and L2 participants take the longest time to complete the 1K Treasure Island adjective sorting task?*

To answer this question, it should be noted that several participants in both the NS and NNS groups commented "Sorting adjectives was difficult" after they completed the task, although they found it difficult to verbalise why they felt like that. The adjectives used in the experiment were chosen from 1K adjectives (e.g., *strong, kind, old, hot, long*), and thus it was not likely that native speakers of English and advanced-level Japanese speakers of English failed to understand their meanings. Furthermore, compared with the complexity of the 1K verbs discussed above, these 1K adjectives did not appear to be semantically complex: Rather they were more straightforward for participants to grasp their meanings since their degrees of polysemous nature were much lower than that of the 1K verbs. However, both NS and NNS groups took statistically significantly longer to complete adjective sorting than verb and noun sorting. Some unique feature of English adjectives, which participants did not readily notice in completing the experiment, seemed to have affected them and produced the distinctive results in task completion time. I argue that it is related to the basic semantic structure of English adjectives which is "given by antonymous pairs, with synonymous adjectives clustering. Thus, every predicative adjective is either a member of an antonymous pair or is similar in meaning to a member of such a pair" (Gross, Fischer & Miller, 1989, p. 92). Many of the adjectives[1] and their semantic relations used in the adjective sorting have this characteristic that Gross et al. (1989) claim, which allows us to explain the difficulty of adjective sorting participants felt and why L1 and L2 participants took the longest to complete it.

Gross et al.'s (1989) claim takes as its basis the findings of Deese (1964, 1965), whose study showed that in word association tests, responses adult English native speakers gave to stimulus adjectives were predominantly antonyms. This includes such stimulus-response pairs as *big-little, soft-hard, tall-short, white-black* and *high-low*, which also had salient features of mutual association (i.e., if a stimulus A produces a response B, then a stimulus B also produces a response A). These distinct results are "fundamental to Deese's assumption that direct antonyms are critical to the representation of English adjectives" (Charles, Reed & Derryberry, 1994, p. 330). Thus, regarding English adjectives in the mental lexicon, their basic semantic relation is likely to be antonyms, and they seem to be highly plausible as cognitive realities. This view was developed by Miller and Fellbaum (1991) with further sophistication of antonym concept and they theorised that "all predicative adjectives have antonyms; those lacking direct antonyms [e.g., *heavy-light*] have indirect antonyms [e.g., *heavy-weightless*], that is, are synonyms of adjectives that have direct antonyms" (p. 211).Thus, every adjective is "either a member of an antonymous pair or is similar in meaning to a member of such a pair" (Gross, Fischer & Miller, 1989, p. 92).

What light, then, can we shed on the adjective sorting task results when we adopt this view of adjectives? What does it matter if their basic semantic structure is given by antonymous pairs, with synonymous adjectives clustering? While we should be cautious about answering this question, we argue that both L1 and L2 participants had trouble with trying to find and often

[1] English adjectives can be categorised into a predicative adjective (which is used before a noun), an attributive adjective (which are used after a verb) and others. The distinction has little to do with the current analysis. Therefore, we do not make a distinction between them in the discussion.

failing to find the "missing links" of antonyms among the tested 1K adjectives in searching for antonymous pairs and synonymous adjective clustering. It was also likely that participants might not even be aware that they were doing the fruitless endeavour unconsciously, but they did take more time to notice that there was no partner to an antonym because of its absence than to find an antonym which was among the tested words. Thus "missing links" of antonyms seem to have affected their sorting behaviour significantly.

Among the adjectives used in the sorting task, there were antonymous pairs and synonymous adjective clustering. They included *recent-last, high-low, near-far, large-small-big-little* and *black-blue-red.* These pairs and clustering seemed to be easy to find and sort, and participants in both groups similarly put them in the same final clusters (sub-clusters) of TIME, SPACE, SIZE and COLOUR (see Tables 7.8 and 7.9 in Chapter 7). On the contrary, the 1K *Treasure Island* adjectives also contained such words as *open, natural, dead, clear, round,* and *afraid*, and these adjectives did not have paired partners that participants might expect to find such as *closed* to *open*, *artificial* to *natural*, *alive* to *dead*, *vague* to *clear*. We cannot measure how frustrated participants felt when they found these "missing links" along with "paired links" among the tested adjectives. However, it can be safely assumed that participants might find it difficult to sort the mixture of adjectives which were composed of "paired links" and "missing links" and often failed to find antonymous pairs (which were absent in the tested words). Eventually, it took them more time to sort the adjectives than to sort the verbs and nouns, which do not have this semantic structure unique to adjectives. When participants were completing the task of "sorting the words into groups according to the meanings you think", they attempted to find something in common in meaning between the tested words. Thus, when they could not find antonymous pairs and synonymous adjectives clustering, participants sorted adjectives of "missing links" into clusters where the words shared some broader meanings with each other. We found evidence for this in the fact that L1 participants made a sub-cluster of positive meanings with *open, natural, clear* and *strong* and L2 participants made a sub-cluster of positive meanings with *ready, successful* and *great.* NS's sub-cluster of SHAPE and SIZE (*round, wide, long, deep, large, big, small* and *little*) and NNS's corresponding sub-cluster (*round, low, long, wide, high, deep, light, little, small, large, big, single* and *alone*) also show that participants in both groups worked hard and took time to find out the semantic relationships between words which had no explicit antonymous pairs and synonymous adjective clustering.

(c) *What is important about the difference of task completion time between the two groups being the smallest in the noun sorting task?*

Both the NS and NNS groups accessed, processed and sorted the 1K *Treasure Island* nouns at almost the same speed, where the mean between-subject (group) difference of task completion time was a mere 0.19 min (NS = 8.41 min and NNS = 8.60 min). The point is that only 1K *Treasure Island* nouns produced statistically significant L1 and L2 differences in aspects of the sorting task results (i.e., mean cluster number, mean number of words per cluster, mean largest cluster participants made, mean distance of each individual participant's dendrogram and difference between L1 and L2 group dendrogram distances). Thus, we concluded that 1K *Treasure Island* nouns can be predictors of L1 and L2 differences in lexical organisation of the mental lexicon. In Chapter 8, we also found that these results were attributed to the fact that English loanwords into the Japanese language played a decisive role and that the nouns used in the noun sorting experiment contained a high proportion of English loanwords (39, 78.0%). Nouns are "often 'over-represented' in early [L1] vocabularies" (Källkvist, 1999, p. 55), and this persistently affects L2 vocabulary learning. By this effect of persistence, English loanwords can be the most difficult L2 lexical items and can hinder developing native-like L2 lexical knowledge and organisation. This is particularly true if the L1 and L2 lexical links and semantic fields were qualitatively different from each other as is the case with high frequency English nouns. Thus, even if the two groups made the smallest difference in task completion time, their sorting behaviour was different from each other. Both groups sorted 1K nouns in almost the same amount of time, but the lexical knowledge and organisation L1 and L2 participants tapped into in task completion appeared to be quite different from each other.

In the 1K *Treasure Island* noun sorting task results, we detected pieces of evidence for the fact that L2 participants developed different lexical links and semantic fields from their L1 counterparts. This included the cases where L2 participants most strongly linked *life* to *rest*, whereas L1 participants sorted *life* into the *sense* cluster containing *sense, wonder, thought* and others and where L2 participants linked *head* to *face*, but L1 participants linked *head* to *shoulder*. These are all related to the persistent effects of English loanwords into the Japanese language in that if L2 lexical items have first been learned as L1 lexical items with "false cognate status" like English loanwords and fossilised into the learner mental lexicon, they seem to be extremely difficult to be re-learned and restructured into native-like knowledge and structure. This issue will be further addressed in the following section.

9.5 Ultimate attainment

We found that there were distinctive differences of lexical organisation in the L1 and L2 mental lexicons. This can be interpreted to mean that it might be difficult for second language learners to reach the ultimate attainment of native-like lexical knowledge and organisation, though not impossible. In this section, we will address this problem. Three issues will be discussed. First, we will address L1 transfer and fossilisation in the development of L2 lexical organisation and address which phase of L2 lexical development plays a role in keeping it from achieving native-like attainment. The discussion will include the interpretation of the present findings while keeping in mind the 'tasks involved in learning the meaning of words' (i.e., labelling, packaging and network-building) (Aitchison, 1994, 2003). Second, we will discuss what can be the barriers against native-like lexical organisation for non-native speakers. The discussion will include gaps of semantic frames (i.e., fields) (Fillmore, 1982; Fillmore & Atkins, 1992) between L1 and L2 mental lexicons and L1 conceptual transfer in L2 lexical acquisition. Third, we will address the relationships between L1 and L2 lexical links and concept links. Kroll and Tokowicz's (2001) Revised Hierarchical Model will be discussed.

9.5.1 L1 transfer and fossilisation in L2 lexical development

In this section, while keeping in mind the findings in the 1K *Treasure Island* noun sorting task results, we will discuss L1 transfer (the influence of the first language) and fossilisation (the permanent cessation of learning in a second language) in L2 vocabulary acquisition (for more on language transfer and fossilisation, see Cook, 1995; Han & Selinker, 1999; Kellerman, 1989; Odlin, 1989, 1993; Selinker, 1972; Selinker & Lakshamanan, 1992).

Aspects of vocabulary acquisition and development can be validly approached by psycholinguistics and connectionist paradigms, as has been done in this project and other mainstream L2 vocabulary acquisition studies which have employed word association tests (e.g., Fitzpatrick, 2006, 2007, 2008; Meara, 1983, 1992, 2006, 2007a, 2007b; Meara & Schur, 2002; Meara & Wolter, 2004; Orita, 2000, 2002a, 2002b; Schur, 2003; Söderman, 1993; Wilks, 1999; Wilks & Meara, 2002; Wolter, 2005). Meara and Wolter (2004) stated that "the really interesting feature of vocabularies is the way that the individual words that make them up interact with each other" (p. 88). Our sorting task experiments have attempted to reveal L1 and L2 differences of lexical knowledge and organisation along these lines. In addition, 'knowing a word' is comprised of inter-related facets to be learned by learners themselves. They are 'form' (which includes pronunciation, spelling and word parts), 'meaning' (which is comprised of concepts, form-meaning connection and associations) and 'use' (which includes grammar, collocations and constraints) (Nation, 2001). While we recognise this nature of inter-relatedness of 'knowing a word', the main realm that the sorting tasks tapped into was meaning. Thus, keeping the sorting task results in mind, it will be worthwhile for us to examine which phase of 'knowing a word' plays a role in affecting the difficulties L2 participants seem to have in attaining native-like lexical knowledge and organisation.

We found that 1K *Treasure Island* nouns can be the predictors that distinguish L1 participants from their L2 counterparts in lexical knowledge and organisation. Distinctive L1 and L2 differences were detected in mean cluster number, words per cluster, mean largest cluster, mean distance of each individual participant's dendrogram and the difference between L1 and L2 group dendrogram distances. Moreover, their group final clusters were qualitatively different from each other. Our examination revealed that a high ratio of loanwords in the 1K *Treasure Island* sorting task (39 words = 78.0%) led participants to produce the marked difference. The key point is that L2 participants maintained and transferred their L1 (Japanese) semantic fields and this effect was too strong to bring about the formation of native-like semantic fields in their L2 mental lexicon. Thus, English loanwords into the Japanese language show that there are areas of L2 lexical knowledge and organisation development that are persistently affected by L1 lexical knowledge. Ultimately relearning or restructuring them into L2 equivalent lexical items and structures is quite difficult and the learner lexical knowledge and structure is fossilised. In L1 lexical development, particularly in its early stages, one has to go through three different but related tasks, those being the labelling task (where one must discover that sequences of sound can be used as names for things), the packaging task (where one must find out which things can be packaged together under one label) and the network-building task (where one must work out how words relate to one another) (for details, see Aitchison, 1994, 2003 and Haastrup & Henriksen, 2000). It is very difficult for the fossilisation that takes place in the L1 labelling phase (as is the case with English loanwords in the Japanese language) to be restructured. It is even more difficult for packaging and network-building to be established. Fossilisation that is attributed to L1 lexical acquisition affects the labelling, packaging and network-building tasks in L2 vocabulary acquisition that L2 learners have to go through. The results showed that it is very hard to develop fossilised lexical knowledge and organisation, which Japanese EFL learners are perhaps not even aware of, into native-like knowledge and organisation.

We should note another problem related to L1 transfer and fossilisation, that is, the difficulty in the packaging task to be faced by L2 learners whose first language contains a large number of loanwords imported from the target language. For the packaging task to be properly carried out, it requires L2 learners to have much authentic input through reading, listening and communication with native-speakers, and the process takes much time and effort. Lafford, Collentine and Karp (2000) pointed out that L2 learners often make naive assumptions about the relationship between their L1 and L2 (e.g., words have exact equivalents between the different languages). When L2 learners notice this is not always the case, they seek for other potential meanings. Thus, Lafford et al. indicate that cognates (loanwords) can be a help to L2 learners, but false cognates cannot. In the case of many English loanwords in the Japanese language, false cognates can be a big hindrance for L2 learners in building native-like packaging of related lexical items. This is because Japanese learners of English as an L2 might often transfer L1 lexical knowledge and packaged semantic fields that are strikingly different from those of the target language. Eventually, L2 learners tend to develop packaged L2 lexical items which are qualitatively different from those of the target language. The findings can be explained by this difficulty L2 learners have to face in the packaging task, together with that in the labelling task discussed above. Needless to say, these problems must have affected the network-building phase in the long run and produced distinctive L1 and L2 differences in the 1K *Treasure Island* sorting task results addressing L1 and L2 lexical organisations.

Considering these facts together, we may hypothesise that L1 words can often be the source of lexical fossilisation and keep L2 learners from establishing native-like L2 lexical organisation, as is the case with English nouns that are first learned as loanwords in Japanese. On the other hand, L1 words that are true cognates (i.e., have the exact same equivalent in the target language regarding aspects of lexical knowledge) will be the L2 words that will be most easily learned and possible to attain native-likeness. Non-loanword status L2 words appear to be much easier for L2 learners to achieve the ultimate attainment of native-like lexical knowledge and organisation than loanwords status L2 words which almost always generate L1 transfer and fossilisation.

9.5.2 Barriers preventing native-like attainment

In this section, we will address factors which seem to hinder L2 participants from developing native-like attainment of lexical knowledge and organisation. Among the plausible factors, three points in particular will be discussed. They are (a) why English loanwords in the Japanese language can be difficult for Japanese learners of English to relearn as L2 lexical items, (b) what additional element of 'knowing a word' plays a role in generating an almost unbridgeable gap between L1 and L2 lexical organisation, and (c) what evidence there is against the learner difficulties in native-like attainment of lexical knowledge and organisation.

First, in an EFL (English as a foreign language) environment, where the dominant language for daily communication is not the target language but instead the first language, it is almost impossible for L2 learners to be aware of the rich meanings of loanwords and restructure them in L2 vocabulary learning. In Chapter 8, we indicated that English loanwords imported into Japanese lexical items are prevalent, accounting for 10.1% of spoken Japanese (Inagaki, 1991), and 80.8% of loanwords imported into Japanese are from English (Tamamura, 1991). As discussed above, the problem is that the lexical knowledge and semantic fields which Japanese EFL learners have developed are often qualitatively different from those of the English language. The persistent effect in L2 learning was found in the qualitative and quantitative differences of the final clusters the NS and NNS groups made, which were reported in the previous chapters. The robustness of L1 transfer, which was evidenced by L1 and L2 differences in the final clusters, seems to have been caused partly by the fact that L2 learners in an EFL environment use only one meaning of a loan word in L1 (Japanese) communication. Moreover, Japanese EFL learners "tend to be familiar only with its colloquial meaning usage in Japanese" (Benthuysen, 2004, p. 171). They have more chances to further strengthen (fossilise) the knowledge of a single meaning of an English loan word than to restructure and relearn it, whereas an L2 lexical item has richer multiple meanings they should learn. Osaka (2005) names this phenomenon 'semantic narrowing' where "words with a number of prominent meanings in English are adopted using only one of these meanings. For example, *channeru* (*channel*) generally means *TV channel* in Japanese but not *waterway*. In this case, the meaning of *channeru* is strictly limited" (p. 163). Thus, even though Japanese EFL learners have chances to encounter loan word status English words in reading, listening and learning the English language even in an EFL environment, they have quite a limited chance to relearn them properly and develop native-like lexical knowledge because of their L1 use and fossilisation. The sorting task results showed that this held true with advanced-level EFL learners who were the Japanese participants in the present sorting task experiments.

Second, and related to the first point, there appears to be an additional element of 'knowing a word' that accounts for creating an almost insurmountable gap between L1 and L2 lexical organisation. Considering the sorting task results, particularly the final clusters the NS and NNS groups produced, the gap is concerned with a semantic 'frame (schema/script)' that underlies "any system of concepts related in such a way that to understand any one of them you have to understand the whole structure in which it fits" (Fillmore, 1982, p. 111). Fillmore and Atkins (1992) explain the concept of semantic frames by giving the example of the RISK frame, which is composed of *risk, danger, peril, hazard* and *venture*, similar words such as *gamble, invest* and *expose* as well as derivatives (e.g., *venturesome, risky, investment* and *perilously*). In this view of cognitive/lexical structure, we should note that the "whole set includes verbs, nouns, adjectives, adverbs, and conventionalized phrases" (Fillmore & Atkins, 1992, pp. 79-80). The sorting task results showed that this semantic frame, which corresponds to a 'semantic field' we have used in this thesis, is psycholinguistically real in the mental lexicons of L1 and L2 speakers. The final clusters identified by our examination represent one of these cognitive/lexical structures (e.g., verbs of THINK, verbs of GET, verbs of GO and verbs of FIGHT). Particularly, the mixed word sorting task revealed that both L1 and L2 participants made final clusters that were made up of words of various word types such as the dimension of nature (*matter, form, clear, nature, air*) and the dimension of quantity (*also, very, lot, nothing, all*) (see Tables 9.1 and 9.2). Thus, it is confirmed that the sorting task results tapped into the semantic frames (i.e., fields) of the L1 and L2 lexical organisations from a cognitive linguistic perspective of 'knowing a word' as well.

What matters is that the semantic frames which the NS and NNS groups made were consistently and qualitatively different from each other across the sorting task results. For example, regarding the 1K *Treasure Island* adjective sorting task, the NS group made five final clusters: 1) positive meanings and emotion, 2) negative meanings of emotions, 3) physical dimensions of degrees, 4) physical conditions of time, space, shape and size and 5) physical dimensions of colours, whereas the NNS group made three final clusters: 1) positive and negative meanings and emotions, 2) dimensions of time, space, degrees, shape and size and 3) physical dimensions of colours. Thus the two groups produced distinctively different final clusters except for the physical dimensions of colours (*black, blue, red*), whose semantic frame is shared between the two lexical organisations. Furthermore, there was also a quantitative difference between the two groups in that the L1 and L2 differences of semantic frames are detected in the number of final clusters, where L1 participants consistently produced a larger number of final clusters than their L2 counterparts. To verify this, refer to Table 9.9, which shows the numbers of final clusters that NS and NNS produced in the 1K *Treasure Island* verb, adjective and noun sorting tasks.

Table 9.9. Word types and number of final clusters

	Treasure verbs	*Treasure* adjectives	*Treasure* nouns
NS ($n = 30$)	8	5	6
NNS ($n = 30$)	6	3	5

Table 9.9 shows that L1 participants always produced a larger number of final clusters than their L2 counterparts across all of the sorting tasks. This is evidence for the claim that native speakers are more aware of lexical items as belonging to distinct sets than non-native speakers are. Thus the L1 and L2 lexical organisations appear to be different from each other both qualitatively and quantitatively. As Bley-Vroman (1989) states, through the knowledge of a language, there is a "full range of subtle intuitions native speakers possess" (p. 51), and this also holds true with lexical knowledge and organisation. Our sorting task data shows that these subtle intuitions are not easy for non-native speakers to acquire.

Third, as a slightly indirect but closely related issue to the barriers against native-like attainment for L2 learners, we will address what evidence there is for the difficulties of native-like attainment of lexical knowledge and organisation. Among them, two issues in particular will be addressed: (a) restructuring of the existing network and (b) semantic constraints of the L1 on L2 vocabulary acquisition.

Wolter (2006) reported that Japanese learners of English quite commonly describe a *room* as *narrow* or *wide*, rather than *small* or *big* even if they are familiar with their meanings. Wolter states that "the process of building syntagmatic connections between words in an L2 appears to be considerably harder than the process for building paradigmatic connections. This is because adding syntagmatic connections will sometimes require restructuring of the existing network, but adding paradigmatic connections will not" (p. 746). His claim is worthwhile to pay attention to in that his analysis reveals that some types of links are more difficult for L2 learners than others regarding restructuring of the existing lexical networks. The type of error in syntagmatic connection he reported is the one that clearly reflects learners' L1 transfer where in the Japanese language, the word for *small* (*chiisai*) does not usually collocate with *room* (*heya*) as often as *narrow* (*semai*). Needless to say, learners' lack of L2 collocational knowledge also plays a significant role in producing this "unlikely, though not wholly unacceptable English collocation" (Wolter, 2006, p. 742). Perhaps this sort of collocation error is not as serious as the issue of lexical fossilisation in lexical organisation discussed in the previous section, which is more deeply rooted in the L2 learner mental lexicon and appears to be far more difficult to overcome. Advanced-level EFL learners, who usually benefit from a large amount of rich input from authentic English even in their EFL environment, are expected to commit lexical errors less frequently in accordance with their L2 proficiency development. In addition, the '*narrow room*' type error is a production-based one and thus can be more easily noticed by the person herself, her teacher or interlocutor. Eventually, advanced-level EFL learners have more chances to notice and correct such an error than a loan word-oriented error which they usually cannot be expected to be aware of.

Another related area that is difficult for L2 learners to overcome is concerned with the semantic constraints of their L1 on L2 vocabulary acquisition. That is, L2 learners persistently maintain their excessive reliance on L1 semantic concepts in L2 production often without noticing it. Odlin (2008) reports an example where Japanese speakers of English seldom chose the verb '*climb*' and instead chose '*go*' in describing a picture of a squirrel climbing down a tree [which was originally reported in Yu (1996)]. Odlin indicated that the avoidance of *climb* among the Japanese "appears to be related to the directionality of the closest translation equivalent, *noboru*, which denotes, according to Yu, motion in only an upward path. Thus, while the heavy reliance on *go* might seem merely attributable to a pattern of overgeneralization independent of the native language, the semantic constraints of the L1 actually help to foster the overreliance" (Odlin, 2008, p. 321). As in the case with the 'narrow/wide room' problem identified above, this sort of learners' unconscious reliance on the L1 semantic concepts in L2 vocabulary acquisition appears to be very difficult for L2 learners to overcome. This reliance on the L1 might be more difficult for L2 learners to eliminate than 'narrow/wide room' type errors in that the latter usually evokes no communication problem when an L2 speaker utters, for example, "A squirrel is going down a tree" instead of "A squirrel is climbing down a tree". The interlocutor would seldom even notice the fact that the L2 learner partially lacks in the lexical knowledge of *climb*. Moreover, the L2 learner herself would have little chance to notice it and expand the semantic richness of the verb in the process of L2 lexical acquisition.

L1 transfer and fossilisation in L2 lexical development we discussed in this section would not explicitly be detected by sorting tasks where the tasks do not require participants to tap into these aspects of lexical knowledge. However, the identified problems, which seem to be difficult to overcome for L2 learners, certainly affect the structuring of L2 lexical organisation and may hinder learners from attaining native-like structure.

9.5.3 Lack of inter-lexical links

In this section, we will address the difficulty of ultimate native-like lexical attainment that is attributed to the lack of inter-lexical links between L1 and L2 lexicons. First, we will examine Kroll and Tokowicz's (2001) Revised Hierarchical Model, a model that most validly depicts the relationships between L1 and L2 lexical links and concept links in the L2 mental lexicon. Second, we will examine a case where the phenomenon of inter-lexical link absence is likely to be seriously problematic for Japanese EFL learners in terms of the ultimate attainment of L2 lexical acquisition.

Kroll and Tokowicz's (2001) Revised Hierarchical Model, which Kroll and her colleagues have developed and improved through a series of experiments (e.g., translation tasks, word naming tasks and reading words aloud tasks), explains the relationships among the L1 lexicon, L2 lexicon and concepts (conceptual knowledge) in the L2 mental lexicon. Kroll and Tokowicz's model had been reproduced in Figure 9.1.

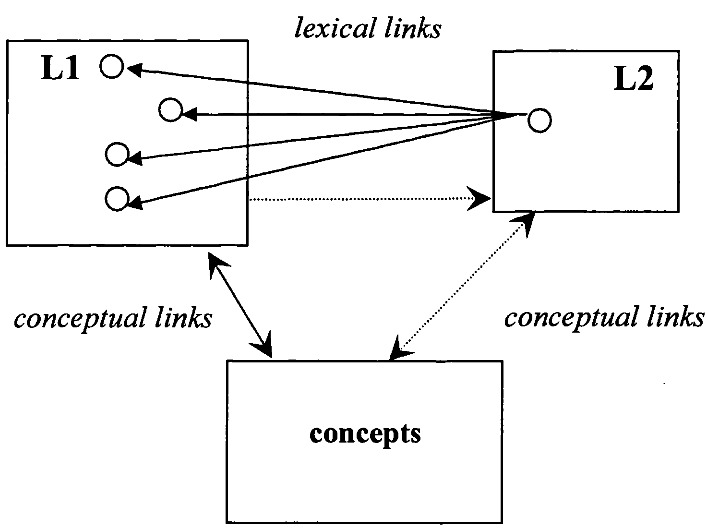

Figure 9.1. Revised Hierarchical Model (Kroll & Stewart, 1994) adapted to show the consequences of the availability of alternative translation equivalents at the lexical level (Kroll & Tokowicz, 2001, p. 62)

This model has four distinctive features. First, it postulates that an L2 lexical item has more than a few translation equivalents. Second, L1 lexical items are more strongly linked to concepts than L2 lexical items are. Third, the size of the L2 lexicon is smaller than its L1 counterpart. Fourth, "L2 words are assumed to be more strongly connected to their corresponding translation equivalents in L1 than the reverse" (Kroll & Tokowicz, 2005, p. 546). In relation to the present research project, the fourth point is important in that it hypothesises the asymmetric organisation of the L2 mental lexicon and presence of translation equivalents between the L1 and L2. Experiments that Kroll and her colleagues carried out have reported evidence for the validity of the Revised Hierarchical Model (see, for example, Habuchi, 2003; Kroll & Dijkstra, 2002; Nakagawa, 2009a, 2009b; Sunderman & Kroll, 2006; Yamashita, 2007). When the L2 mental lexicon is tapped into by lexical decision task-based experiments and the structure is examined by the results of reaction time, it is highly plausible that the model validly explains an essential part of the L2 lexicon.

However, we might need to give an additional explanation to the model when we reconsider it while keeping the sorting task results in mind. That is, the L2 mental lexicon is structured in a more complicated way than the model predicts in that there are lexical items which do not always have translation equivalents between the L1 and L2. We confirmed that English loanwords into the Japanese language have developed semantic frames (i.e., fields) in the lexical organisation of Japanese EFL learners that are qualitatively different from those of native speakers. The frames have been fossilised in their mental lexicon and prevent it from developing into a native-like structure. Conceptual transfer and over-reliance on L1 semantic concepts also function as barriers to attain native-like lexical knowledge and organisation. Thus, given the sorting task results as a whole, the L2 mental lexicon contains more facets of L1 and L2 lexical relationships than the Revised Hierarchical Model postulates. Needless to say, this discussion is from the view point of the present sorting task approach to the L2 mental lexicon and does not intend to argue against the model but to confirm the existence of other essential facets which are not explicitly addressed by the model. The L2 mental lexicon is composed of many lexical links between the L1 and L2 and some of them have translation equivalents and others do not. Therefore finally we will discuss what is probably the most difficult aspect for L2 learners to reach a native-like attainment in lexical knowledge and organisation. That is filling the gap of inter-lexical links between the L1 and L2.

Lack of inter-lexical links between L1 and L2 is concerned with the fact that languages often entail meanings that are unique to themselves, whereas they may also have common, core meanings that are shared between two or more languages. When L2 learners are learning the lexicons of the target language, these vacant parts that exist between the two lexicons appear to be the ones that are the least perceivable and attainable for them. Hara (2004) examined this cross-linguistic semantic lack between Japanese and English while making a comparison of the core and expanded meanings of the term '*eye*' in the two languages. Table 9.10 shows the shared and unshared meanings between the two languages that Hara identified.

Table 9.10. Presence of shared meanings of '*eye*' in Japanese and English

	Entailed Meaning	Japanese '*me*'(eye)	English '*eye*'
(a)	Either of the two organs on the face that one sees with (e.g., *with the naked eye*)	√	√
(b)	The ability to see (e.g., *the surgeon has a good eye*)	√	√
(c)	One's judgment or opinions (e.g., *in the eyes of the law*)	√	√
(d)	Something lying in the centre (e.g., *the eye of a typhoon*)	√	√
(e)	The *eye* of a needle		√
(f)	An *eye* on a potato		√
(g)	A hook and *eye*		√
(h)	An *eye* of the camera		√
(i)	Divisions of a scale	√	
(j)	Bad experience	√	
(k)	The '*teeth*' of a saw	√	

Note. This data is based on Hara (2004). The symbol √ shows that the meaning is present in the language concerned.

Table 9.10 shows that Japanese and English share at least four core meanings of '*eye*' (from (a) to (d) in the table) and those are the meanings that Japanese EFL learners would usually have no difficulty to master because of the direct correspondence between the two languages. Meanwhile, '*eye*' also has extended meanings that are unique to the English language (from (e) to (h) in the table) and their meanings are non-existent in the Japanese '*me* (*eye*)'. Presumably a Japanese EFL learner would not even be conscious of these lacunae unless she received a large amount of rich L2 input that contained these examples of '*eye*' through reading, listening and communication with native speakers of English. What matters is that many high frequency English words have these extended meanings that are non-existent in Japanese and these can be the most unnoticeable and unattainable facets of the L2 lexical items for L2 learners. There are also examples of extended meanings which the Japanese language carries but the English language does not regarding 'eye' (from (i) to (k) in Table 9.11). This might cause another problem if an L2 learner expects these L2 equivalents to be usable as is in the English language. The L2 learner won't find them in English and must then search for another word or phrase to get their meaning across. These are also problems for an EFL learner in achieving ultimate lexical attainment of the target language.

In summary, EFL learners appear to face many types of difficulties in terms of ultimate attainment of L2 lexical knowledge and organisation. It should be noted that easy words, whose form-meaning connections (at least the most frequent, salient ones) L2 learners have learned in the early stages of their vocabulary learning, are actually the most difficult words for them to master fully. These words are comprised of a few core meanings plus several (or sometimes many) extended meanings, and the latter are often not even noticed by L2 learners in the advanced stages of vocabulary learning. It is highly likely that even advanced-level speakers of English as an L2 might have some facets still to be learned and these facets are not easily learned, which can be barriers against attaining native-like lexical knowledge and organisation.

9.6 Conclusion

In this chapter, we confirmed that the methodological improvements we made to the sorting tasks worked well and led them to become reliable in tapping into the differences in L1 and L2 lexical organisations. Task sensitivity was boosted by selecting the words used in the experiments from a cohesive passage in *Treasure Island* and by directing participants to complete the task as quickly as they could. Our analyses of the cluster number participants made revealed that both the NS and NNS groups had non-significant between-subject variability but that they maintained within-subject consistency across the sorting tasks (except in the 1K *Treasure Island* verb sorting task). There was also a very high correlation between the adjective and noun sorting tasks ($r = 0.937, p < 0.001$) for L1 participants. It was found that L2 participants produced the biggest task completion time variability in completing the 1K *Treasure Island* verb sorting task and the result was attributed to verb complexity in syntactic behaviour and the polysemous nature of verbs over other word classes. Both L1 and L2 participants took the longest time in completing the 1K *Treasure Island* adjective sorting task. We argued that the result was attributed to the basic semantic structure of English adjectives which are characterised by antonymous pairs with synonymous adjectives clustering and that participants took time in searching for and failing to find "missing links" among the tested adjectives. Finally, we discussed the issue of the ultimate attainment of native-like lexical knowledge and organisation. We indicated that easy words are in actuality the most difficult words for L2 learners to master because of the decisive role of L1 transfer, fossilisation and lack of inter-lexical links between the L1 and L2.

There remains one question: Is it possible for L2 learners to achieve ultimate native-like attainment of lexical knowledge and organisation? For the moment, we have no definite answer to it because there are many complicated variables involved in addressing the question including the teachability, quantity and quality of L2 input and learner awareness of the attainment problems. Judging from the data and analyses of the present research project, there does not seem to be much promise for teachability. Teaching partial aspects of L2 lexical knowledge in classrooms might be feasible to some extent but teaching L2 lexical organisation is not. Even if L2 learners learned an L1 "expert net" of a limited number of L2 words, it would not give much significant impact on the overall structure of their L2 lexical organisation. More importantly, this type of learning has more to do with memorisation than the restructuring of lexical organisation (see Chapter 2 for the limitations of the instruction using an L1 "expert net" attempted by Sánchez (2004)). Moreover, realistically, English teachers cannot teach all pieces of L2 lexical knowledge and organisation nor can L2 learners learn them all. Classroom vocabulary instruction should be directed towards "the high-frequency words of the language. Where learners are going on to academic study this would also include the *Academic Word List* vocabulary" (Nation, 2001, p. 97). Advanced L2 learners have already gone through this phase of learning and the rich facets of L2 lexis should be learned mainly for themselves in some way or other. Thus, the point is whether they can become conscious of these unperceived facets through reading, listening and communication with native speakers of English. Is it possible for EFL learners to achieve native-like lexical knowledge and organisation while staying in an EFL environment (e.g., their mother land)? Or is it necessary to spend several months or years in an English-speaking country? To the best of our knowledge, we have no research that explicitly addresses these issues.

In regard to the present project, it should be noted that word association tests have "little to say at the moment about how to facilitate this organizational restructuring through teaching" (Schmitt, 2000, p. 42). This is also true with sorting tasks, which are a psycholinguistic approach to the mental lexicon in line with word association tests. Thus, the identified issues in L2 vocabulary acquisition above should be included in the research agenda that will be done in future studies of lexical knowledge and organisation. As Haastrup and Henriksen (2000) did, a longitudinal study needs to be done that collects data of EFL learners regarding their changes in semantic clustering behaviours, the degree of L1 transfer and fossilisation and awareness of absence of inter-lexical links between L1 and L2. Haastrup and Henriksen reported that network building is an extremely slow and gradual process, placing the basis of their claim on a three-year case study of Danish novice EFL learners. Thus, it might be necessary to design a longer longitudinal study to detect marked changes in L2 lexical knowledge and organisation. Meara (1996) states that vocabulary size is the most crucial dimension for novice and intermediate L2 learners to attend to. Once they have reached a threshold (e.g., five or six thousand words), "vocabulary size per se seems to become less important" (p. 45). This implies that other rich dimensions of L2 vocabulary acquisition would also be important for L2 learners once they have reached this threshold level. They include vocabulary depth, lexical fluency and organisation, only the last of which we have addressed in this thesis project. In view of the ultimate attainment of native-like lexical knowledge and organisation, it is worthwhile to address the newly identified issues above in an attempt to shed new light on the structure and restructuring plausibility of the L2 mental lexicon.

Chapter 10: Conclusion

This thesis has attempted to reveal the organisation of the L2 mental lexicon through sorting tasks, a method Haastrup and Henriksen (2000) pioneered. In particular, it was addressed whether the L1 and L2 lexicons are different from each other quantitatively and qualitatively in the clustering behaviours of the NS and NNS groups. The experiments carried out showed that the differences were usually quantitatively subtle, such as in mean cluster number, words per cluster and largest cluster. Simple comparisons of these variables often failed to detect underlying L1 and L2 differences in lexical organisation. As Wilks (1999) predicted, the quantitative differences of the L1 and L2 lexical organisations are more subtle than one would expect. However, the sorting task results consistently showed that the L2 lexical organisation was different from the L1 lexical organisation when the two organisations were analysed by means of cluster analysis. Thus it may be said that the two mental lexicons are qualitatively different from each other, even though they may not be quantitatively different from each other. It is highly plausible that the L2 mental lexicon has developed lexical networks which are on the surface similar to the L1 mental lexicon, when in fact the L2 mental lexicon has established a different organisational structure from its L1 counterpart. As Wilks and Meara (2002) predicted, "two networks with the same density could in fact be quite differently arranged" (p. 319).

It was found that nouns can be the predictors of L1 and L2 differences in all the tested variables of lexical organisation using sorting tasks. The NNS group was distinctively and quantitatively different from the NS group in mean cluster number, words per cluster and largest cluster as well as the mean individual dendrogram and group dendrogram distances. Naturally, the two lexicons were qualitatively different from each other as well. This is firm evidence for Meara and Schur (2002), who reported L1 and L2 differences in these aspects of lexical organisation. However, this is a very unexpected finding since, in L2 vocabulary acquisition, learning nouns are usually not a source of serious difficulty for learners compared to other word classes. Actually, participants in the verb sorting task found it difficult to sort them because of their syntactic complexities and polysemous nature. What matters most is that the degree of difficulty in L2 vocabulary learning does not automatically lead to learner difficulty in attaining native-like L2 lexical organisation. Nouns appear to be easier for Japanese EFL learners to acquire in that the Japanese language has a significant number of nouns that have originated from English loanwords. However, when L2 lexical items are first learned as L1 lexical items with "false cognate status", it is extremely difficult to re-learn and restructure them into native-like L2 knowledge and organisation. This thesis revealed that English nouns fall into this realm of difficulty in developing L2 lexical organisation, at least as appears to be the case for Japanese EFL learners.

In addition, it was found that the intra-lexical features of tested words affected task completion time in a significant way. Its manifestation was different from one word class to another in the experiments. L2 participants produced the biggest task completion time variability in the verb sorting task, both L1 and L2 participants took the longest time to complete the adjective sorting task and the difference of task completion time between the two groups was the smallest in the noun sorting task. These findings suggest that lexical networks in the mental lexicon are established differently from one word type to another and participants activate and process their lexical knowledge and organisation differently. As a result, participants took a different amount of time to complete each sorting task, depending on which word type they were sorting. Thus, task completion time, when compared among sorting tasks, can offer evidence for underlying differences in lexical organisation of tested words in the L1 and L2 mental lexicons.

As is the case with most studies, this thesis has some limitations. One of them is that the L2 participants in this study were only advanced-level Japanese speakers of English, although the participants varied from experiment to experiment. Thus, the findings we made need to be further tested using NNS groups whose mother tongues are non-Japanese, including languages that are both similar and dissimilar to the English language. A second limitation is that this project tested only 1K high frequency English words and a cohesive passage of *Treasure Island* was adopted in the word selection process for the final three revised sorting tasks (which were boosted in task sensitivity). It would be advisable to use less frequent words in future experiments to confirm the validity of the findings this thesis has made. A third limitation, which raises new questions, is related to the finding that English loanword nouns in the Japanese language played a role in hindering L2 participants from developing native-like L2 lexical organisation. In Japanese, these nouns have often developed different semantic fields from those of English, and they have been fossilised in the mental lexicons of Japanese EFL learners, preventing them from being re-learned and restructured. Thus, new questions arise: Is it possible for L2 learners to attain native-like L2 lexical knowledge and organisation for the L2 words which are free from the effect of false cognates such as English loanwords in Japanese? Is there any L2 lexical knowledge and organisation that can be developed, independent of L1 lexical knowledge and conceptual knowledge? That is, is it possible for L2 learners to establish native-like L2 lexical organisation for the L2 lexical items that have no L1 translation equivalents and concepts? These questions need to be addressed before attempting to make generalisations about the findings this thesis has made.

Thus, this thesis has also revealed new issues that will be worthwhile to address in future research in order to deepen our understanding of L1 and L2 differences in lexical organisation. It is hoped that the methodological innovations introduced in this thesis as well as some of the findings this thesis project has reported will contribute to a better understanding of aspects of the L1 and L2 mental lexicon that have yet to be discovered.

Bibliography

Aitchison, J. (1987). *Words in the mind.* Oxford: Blackwell.

Aitchison, J. (1994). *Words in the mind* (2nd ed.). Oxford: Blackwell.

Aitchison, J. (2003). *Words in the mind* (3rd ed.). Oxford: Blackwell.

Aizawa, K. (2003a). Donoyoo ni shite goi wo minitsukete irunoka: Juyoo goi no teichaku kara happyoo goi e [How learners acquire vocabulary: From the acquisition of receptive vocabulary to productive vocabulary]. *Eigo Kyooiku, 52*(7), 17-20.

Aizawa, K. (2003b). Tango wa doo yatte oboeteikunoka [How words are learned]. In M. Mochizuki, K. Aizawa & Y. Tono (Eds.), *Eigo goi no shidoo manuaru* [Teaching manual for English vocabulary] (pp. 71-106). Tokyo: Taishukan.

Ando, K. (1997). Katakana eigo to eigokyooiku [*Katakana* (A Japanese system of phonograms) English and English Education] (2). *Nagoya Keizai University Jimmonkagaku-Ronshuu, 51*, 1-23.

Anglin, J. M. (1970). *The growth of word meaning.* Cambridge, MA: The MIT Press.

Appel, R. (1989). Bilingualism and cognitive-linguistic development: Evidence from a word association task and a sorting task. *Journal of Multilingual and Multicultural Development, 10*(3), 183-196.

Bendix, E. H. (1966). *Componential analysis of general vocabulary: The semantic structure of a set of verbs in English, Hindi, and Japanese.* The Hague: Mouton.

Bensoussan, M., & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading, 7,* 15-32.

Benthuysen, R. V. (2004). Japanese EFL students' awareness of English loanword origins. *Bunkyoo Gakuin Daigaku Gaikokugogakubu-Bunkyoo Gakuin Tanki Daigaku Kiyoo, 4,* 169-174.

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development, 70*(3), 636-644.

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language, 68*(4), 706-755.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 41-68). Cambridge: Cambridge University Press.

Bogaards, J. (1994). *Le vocabulaire dans l'apprentissage des languages étrangères.* Paris: Didier.

Borges, M. A., & Sawyers, B. K. (1974). Common verbal quantifiers: Usage and interpretation. *Journal of Experimental Psychology, 102,* 335-338.

Branaghan, R. J. (1990). Pathfinder networks and multidimensional spaces: Relative strengths in representing strong associates. In R. W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization* (pp. 111-120), Norwood, NJ: Ablex.

Brown, C. (1993). Factors affecting the acquisition of vocabulary: Frequency and saliency of words. In T. Huckin, M. Haynes & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 263-284). Norwood, NJ: Ablex.

Burton, M. (1972). Semantic dimensions of occupation names. In A. Romney, R. Shepard & S. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences* (pp. 55-71). New York: Seminar Press.

Burton, M. L. (1975). Dissimilarity measures for unconstrained sorting data. *Multivariate Behavioral Research. 10,* 409 - 424.

Carter, R. (1987). *Vocabulary: Applied linguistic perspectives.* London: Routledge.

Chaffin, R., & Herrmann, D. J. (1984). The similarity and diversity of semantic relations. *Memory & Cognition, 12*(2), 134-141.

Chapman, R. L. (Ed.). (1992). *Roget's international thesaurus* (5th ed.). New York: Harper Perennial.

Charles, W. G., Reed, M. A., & Derryberry, D. (1994). Conceptual and associative processing in antonymy and synonymy. *Applied Psycholinguistics, 15,* 329-354.

Chen, C. (2003). *Mapping scientific frontiers: The quest for knowledge visualization.* London: Springer.

Chronicle, E. P., MacGregor, J. N., Lee, M. D., Ormerod, T. C., & Hughes, P. (2008). Individual differences in optimization problem solving: Reconciling conflicting results. *The Journal of Problem Solving, 2*(1), 41-49.

Cook, V. (1995). Multi-competence and effects of age: The age factor in second language acquisition. In D. M. Singleton & Z. Lengyel (Eds.), *The age factor in second language acquisition: A critical look at the critical period hypothesis* (pp. 51-66). Clevedon: Multilingual Matters.

Cooke, N. J. (1990). Using Pathfinder as a knowledge elicitation tool: Link interpretation. In R. W. Schvaneveldt (Ed.) *Pathfinder associative networks: Studies in knowledge organization* (pp. 227-239), Norwood, NJ: Ablex.

Coxon, A. P. (1999). *Sorting data: Collection and analysis.* Thousand Oaks, CA: Sage.

Dearholt, D. W., & Schvaneveldt, R. W. (1990). Properties of pathfinder networks. In R. W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization* (pp.1-30), Norwood, NJ: Ablex.

Deese, J. (1964). The associative structure of some common English adjectives. *Journal of Verbal Learning and Verbal Behavior, 3,* 347-357.

Deese, J. (1965). *The structure of associations in language and thought.* Baltimore, MD: The

John Hopkins Press.

de Groot, A.M.B. (1993). Word-type effects in bilingual processing tasks: Support for a mixed representational system. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 27-51). Amsterdam: John Benjamins.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22,* 499-533.

Dietrich, R. (1989). Nouns and verbs in the learner's lexicon. In H. W. Dechert (Ed.), *Current trends in European second language acquisition research* (pp. 13-22). Clevedon, Avon: Multilingual Matters.

Dijkstra, T. (2005). Bilingual visual recognition and lexical access. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism* (pp. 179-201). New York: Oxford University Press.

Dörnyei, Z. (2006). Individual differences in second language acquisition. *AILA Review, 19,* 42-68.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589-630). Oxford: Blackwell.

Ellis, R. (1994). A theory of instructed second language acquisition. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 79-114). London: Academic Press.

Ellis, R. (1997). *SLA research and language teaching.* Oxford: Oxford University Press.

Espe, H. (1985). A cross-cultural investigation of the graphic differential. *Journal of Psycholinguistic Research, 14*(1), 97-111.

Faber, P., & Pérez, C. (1993). Image schemata and light: A study of contrastive lexical domains in English and Spanish. *Atlantis, 15*(1-2), 117-134.

Ferrer i Cancho, R., & Solé, R. V. (2001). The small world of human language. *Proceedings of the Royal Society of London Series B, 268,* 2261-2266.

Field, J. (2003). *Psycholinguistics: A resource book for students.* London: Routledge.

Fillenbaum, S., & Rapoport, A. (1971). *Structures in the subjective lexicon.* New York: Academic Press.

Fillmore, C. J. (1982). Frame semantics. In Linguistics Society of Korea (Ed.), *Linguistics in the morning calm,* (pp. 111-137). Seoul: Hanshin.

Fillmore, C. J., & Atkins, B. T. (1992).Toward a frame-based lexicon: The semantics of RISK and its neighbors. In A. Lehrer & E. F. Kittay (Eds.), *Frames, fields, and contrasts: New essays in semantic and lexical organisation.* Hillsdale, NJ: Lawrence Erlbaum.

Fitzpatrick, T. (2006). Habits and rabbits: Word associations and the L2 lexicon. *EUROSLA Yearbook, 6,* 121-145.

Fitzpatrick, T. (2007). Word association patterns: Unpacking the assumptions. *International Journal of Applied Linguistics, 17*(3), 319-331.

Fitzpatrick, T. (2008, December). *Using word associations to explore vocabulary knowledge.* Paper presented at the 1st JACET Joint Research Forum on English Vocabulary and Reading. Kwansei Gakuin University, Nishinomiya, Japan.

Folarin, B. A. (1989). An investigation of children's word association. *Psychology: A Journal of Human Behavior, 26*(2, 3), 60-64.

Francis, W. S. (2005). Bilingual semantic and conceptual representation. In J. F. Kroll & A.M.B. de Groot (Eds.) *Handbook of bilingualism: Psycholinguistic approaches* (pp. 251-267). New York: Oxford University Press.

Friendly, M., & Glucksberg, S. (1970). On description of subcultural lexicons: A multidimensional approach. *Journal of Personality and Social Psychology, 14,* 55-65.

Furuie, T. (2003).Goi wo oboeyasuku suru mono wa nani ka? [What can help make vocabulary learning easier?]. *Eigo Kyooiku, 52*(7), 11-13.

Garman, M. (1990). *Psycholinguistics.* Cambridge: Cambridge University Press.

Garskof, B. E., & Houston, J. P. (1963). Measurement of verbal relatedness: An idiographic approach. *Psychological Review, 70,* 277-288.

Gass, S. (1988). Integrating research areas: A framework for second language studies. *Applied Linguistics, 9,* 198-216.

Gass, S. M., & L. Selinker. (1994). *Second language acquisition.* Hillsdale, NJ: Lawrence Erlbaum.

Gass, S. M., & L. Selinker. (2001). *Second language acquisition* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and Brain Theory, 4,* 161-178.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj (Ed.), *Language development: Vol. 2. Language, thought and culture* (pp. 301-334). Hillsdale, NJ: Lawrence Erlbaum.

Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge. *Journal of Educational Psychology, 83*(1), 88-96.

Gougenheim, G., Michéa, R., Rivenc, P., & Sauvageot, A. (1956). *L'élaboration du Français élémentaire*. Paris: Didier.

Gross, D., Fischer, U., & Miller, G. A. (1989). The organization of adjectival meanings. *Journal of Memory and Language, 28,* 92-106.

Gross, D., & Miller, K. J. (1990). Adjectives in WordNET. *International Journal of Lexicography, 3*(4), 265-277.

Guasti, M. T. (2002). *Language acquisition: The growth of grammar.* Cambridge, MA: The

MIT Press.

Haastrup, K., & Henriksen, B. (1998). Vocabulary acquisition: From partial to precise comprehension. In K. Haastrup & Å. Viberg (Eds.), *Travaux de l'Institut de Linguistique de Lund: Vol. 38. Perspectives on lexical acquisition in a second language* (pp. 97–114). Lund, Sweden: Lund University Press.

Haastrup, K., & Henriksen, B. (2000). Vocabulary acquisition: Acquiring depth of knowledge through network building. *International Journal of Applied Linguistics, 10*(2), 221-240.

Habuchi, Y. (2003). Word processing in cross-language translation between Japanese and English by advanced second-language learners: A test of the Revised Hierarchical Model. *Japanese Journal of Educational Psychology, 51,* 65-75.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English.* London: Longman.

Han, Z., & Selinker, L. (1999). Error resistance: Towards an empirical pedagogy. *Language Teaching Research, 3*(3), 248-275.

Hara, N. (2004). *Nichieigo ni okeru Shintaibui Hyoogen ni tsuite:* Me *(Eye) to* eye *no Hikakubunseki* [Body-Part Related Expressions in Japanese and English: A Comparative Analysis of *me* (*eye*) and *eye*]. Unpublished undergraduate graduation thesis, the Prefectural University of Kumamoto, Kumamoto, Japan.

Harley, B. (1995). The lexicon in second language research. In B. Harley (Ed.), *Lexical issues in language learning* (pp. 1-28). Amsterdam: John Benjamins.

Hatch, E., & Brown, C. (1995). *Vocabulary, semantics, and language education.* Cambridge: Cambridge University Press.

Henriksen, B. (1996, August). Semantisation — A key process for vocabulary learning and use. Paper presented at the AILA conference. Jyväskylä, Finland.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21,* 303-317.

Henriksen, B. (2003, September). *Operationalising the construct of organisational structure of the L2 lexicon.* Paper presented at EUROSLA 2003. Abstract retrieved 14 June, 2009, from http://www. hw.ac.uk/langWWW/eurosla/abstracts/bhenriksen.htm

Henriksen, B., & Haastrup, K. (1998). Describing learners' lexical competence across tasks and over time: A focus on research design. In K. Haastrup & Å. Viberg (Eds.), *Travaux de l'Institut de Linguistique de Lund: Vol. 38. Perspectives on Lexical Acquisition in a Second Language* (pp. 61–95). Lund, Sweden: Lund University Press.

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*(2), 689-696.

Hirshman, E. C., & Wallendorf, M. R. (1982). Free-response and card-sort techniques for assessing cognitive content: Two studies concerning their stability, validity and utility. *Perceptual and Motor Skills, 54,* 1095-1110.

Hyltenstam, K., & Abrahamsson, N. (2000). Who can become native-like in a second language? All, some or none? *Studia Linguistica, 54*(2), 150-166.

Ikemura, D., & Kadota, S. (2003). Go no imi hyooshoo e no akusesu [Access to the representations of word meanings]. In S. Kadota (Ed.), *Eigo no mentaru rekishikon* [The English mental lexicon] (pp. 99-122). Tokyo: Shohakusha.

Inagaki, S. (1991). Gaigaigo hyooki no kijun to kanyoo [Criteria for and usages in the notation of loan words in Japanese]. *Nihongo Kyooiku, 74,* 60-72.

Jaworska, E. (1998). Cohesion. In K. Johnson & H. Johnson (Eds.), *The encyclopedic dictionary of applied linguistics* (pp. 55-57). Oxford: Blackwell.

JACET (The Japan Association of College English Teachers) Basic Vocabulary Revision Committee. (2003). *JACET List of 8000 Basic Words.* Tokyo: The Japan Association of College English Teachers.

Johnson, E. J., & Tversky, A. (1984). Representations of perceptions of risks. *Journal of Experimental Psychology: General, 113*(1), 55-70.

Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge.* Hillsdale, NJ: Lawrence Erlbaum.

Kadota, S. (2003). Goi nettowaaku to daiichi/daini gengo no mentaru rekishikon [Lexical networks and the L1 and L2 mental lexicons]. In S. Kadota (Ed.), *Eigo no mentaru rekishikon* [The English mental lexicon] (pp. 217-243). Tokyo: Shohakusha.

Kamada, T, & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters, 31,* 7-15.

Kellerman, E. (1989). The imperfect conditional. In K. Hyltenstam & L. K. Obler (Eds.), *Bilingualism across the lifespan: Aspects of acquisition, maturity, and loss* (pp. 87-115). Cambridge: Cambridge University Press.

Koojien [The Encyclopaedia of the Japanese Language] (6th ed.). 2006. Tokyo: Iwanami Shoten.

Kroll, J.F. (1993). Accessing conceptual representations for words in a second language. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 53-81). Amsterdam: John Benjamins.

Kroll, J.F., & de Groot, A.M.B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In A.M.B. de Groot & J.F. Kroll (Eds.), *Tutorials in bilingualism* (pp.169-199). Mahwah, NJ: Lawrence Erlbaum.

Kroll, J. F., & Dijkstra, T. (2002). The bilingual lexicon. In R. B. Kaplan (Ed.), *The Oxford* *handbook of applied linguistics* (pp. 301-321). Oxford: Oxford University Press.

Kroll, J. F., & Potter, M. C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *Journal of Verbal Learning and Verbal Behavior, 23,* 39-66.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33,* 149-174.

Kroll, J. F., & N. Tokowicz. (2001). The development of conceptual representation for words in a second language. In J. L. Nicol (Ed.) *One mind, two languages* (pp. 48-71). Malden, MA: Blackwell.

Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism* (pp. 531-553). New York: Oxford University Press.

Kruse, H., Pankhurst, J., & Sharwood Smith, M. (1987). A multiple word association probe in second language acquisition research. *Studies in Second Language Acquisition, 9,* 141-154.

Kunihiro, T. (1999). Imi to koozoo no hikaku [A comparison of English and Japanese lexical organisations]. In T. Kunihiro (Ed.), *Imi to goi* [Meanings and vocabulary] (pp. 15-52). Tokyo: Taishukan.

Källkvist, M. (1998). Lexical infelicity in English: The case of nouns and verbs. In K. Haastrup & A. Viberg (Eds.), *Perspectives on lexical acquisition in second language* (pp. 149-174). Lund: Lund University Press.

Källkvist, M. (1999). *Form-class and task-type effects in learner English: A study of advanced Swedish learners.* Lund: Lund University Press.

Lafford, B. A., Collentine, J., & Karp, A. (2000). The Acquisition of lexical meaning by second language learners: An analysis of general research trends with evidence from Spanish. Retrieved 7 May, 2009, from http://jan.ucc.nau.edu/~jgc/research/vocabstate/

Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics.* Oxford: Oxford University Press.

Laufer, B. (1997). What's in a word that makes it hard or easy: Some intralexical factors that affect the learning of words. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 140-155). Cambridge: Cambridge University Press.

Lennon, P. (1996). Getting 'easy' verbs wrong at the advanced level. *International Review of Applied Linguistics in Language Teaching, 34*(1), 23-36.

Lewis, M. (1997). *Implementing the lexical approach.* Hove: Language Teaching Publications.

Long, M. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12,* 251-285.

Lyons, J. (1968). *Introduction to theoretical linguistics.* London: Cambridge University Press.

Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design.* Mahwah, NJ: Lawrence Erlbaum.

MacWhinney, B. (1997). Second language acquisition and the Competition Model. In J. Kroll & A. De Groot (Eds.), *Tutorials in bilingualism* (pp. 113-142). Mahwah, NJ: Lawrence Erlbaum.

MATLAB (Matrix Laboratory) (Version 7.5.0.342 (R2007b) [Computer software]. (The MathWorks, Inc., 2007)

McCarthy, M. (1990). *Vocabulary.* Oxford: Oxford University Press.

Meara, P. (1978). Learners' word associations in French. *Interlanguage Studies Bulletin, 2,* 192-211.

Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching and Linguistics: Abstracts, 13,* 221-246.

Meara, P. (1983). Word associations in a foreign language. Retrieved 23 June 2009, from http://www.lognostics.co.uk/vlibrary/meara1983.pdf

Meara, P. (1992). Network structures and vocabulary acquisition in a foreign language. In P. Arnaund & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 62-70), London: Macmillan.

Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjær & J. Williams (Eds.), *Performance and competence in second language acquisition* (pp. 33-53). Cambridge: Cambridge University Press.

Meara, P. (1997). Towards a new approach to modelling vocabulary acquisition. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 109-121). Cambridge: Cambridge University Press.

Meara, P. (2001, November). *Q_Lex: An alternative approach to vocabulary knowledge.* Paper presented at JALT/PAC 3. Kitakyushu, Japan.

Meara, P. (2002). The rediscovery of vocabulary. *Second Language Research, 18(4),* 393-407.

Meara, P. (2004a). Psycholinguistics : A resource book for students [Review of the book *Psycholinguistics: A resource book for students*]. *System, 32,* 122-125.

Meara, P. (2004b). Modelling vocabulary loss. *Applied Linguistics, 25*(2), 137-155.

Meara, P. (2005). Lexical frequency profiles: A Monte Carlo analysis. *Applied Linguistics, 26*(1), 32-47.

Meara, P. (2006). Emergent properties of multilingual lexicons. *Applied Linguistics, 27*(4), 620-644.

Meara, P. (2007a). Simulating word associations in an L2: The effects of structural complexity. *Language Forum, 33,* 13-31.

Meara, P. (2007b). Growing a vocabulary. *EUROSLA Yearbook 7,* 49-65.

Meara, P., & Schur, E. (2002). Random association networks: A baseline measure of lexical complexity. *British Studies in Applied Linguistics, 17,* 169-182.

Meara, P., & Wolter. B. (2004). V_Links: Beyond Vocabulary Depth. *Angles on the English-Speaking World, 4,* 85-97.

Miller, G. A. (1967). Psycholinguistic approaches to the study of communication. In D. L. Arm (Ed.), *Journeys in science: Small steps—great strides* (pp.22-73). Albuquerque, New Mexico: The University of New Mexico Press.

Miller, G. (1969). A psychological method to investigate verbal concepts. *Journal of Mathematical Psychology, 6,* 169-191.

Miller, G. A. (1972). Lexical memory. *Proceedings of the American Philosophical Society, 116*(2), 140-144.

Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes, 6*(1), 1-28.

Miller, G. A., & Fellbaum, C. (1991). Semantic networks of English. *Cognition, 41,* 197-229.

Mochizuki, M. (2002). Exploration of two aspects of vocabulary knowledge: Paradigmatic and collocational. *Annual Review of English Language Education in Japan, 13,* 121-129.

Moxey, L. M., & Sanford, A. J. (1992). Context effects and the communicative functions of quantifiers: Implications for their uses in attitude research. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 279-296). New York: Springer Verlag.

Moxey, L. M., & Sanford, A. J. (1993). Prior expectation and the interpretation of natural language quantifiers. *European Journal of Cognitive Psychology, 5,* 73-91.

Nakagawa, C. (2009a). Examination of the developmental hypothesis on the Revised Hierarchical Model. *Annual Review of English Language Education in Japan, 20,* 121-130.

Nakagawa, C. (2009b). Effects of encoding processes on L2 lexical network activation. *JACET Journal, 48,* 39-52.

Nakanishi, Y. (2003a). Tango no shori to kioku: Eitango risuto no jiyuusaisei kadai ni yoru kenshoo [Processing words and memory: An examination by means of a free-recall task of English words]. In S. Kadota (Ed.), *Eigo no mentaru rekishikon* [The English mental lexicon] (pp. 173-196). Tokyo: Shohakusya.

Nakanishi Y. (2003b). Goi chishiki to sono sokutei [Vocabulary knowledge and its measurement]. In S. Kadota (Ed.), *Eigo no mentaru rekishikon* [The English mental lexicon] (pp. 32-44). Tokyo: Shohakusha.

Nation, P. (1986). *Vocabulary lists: Words, affixes and stems.* Occasional Publication No. 12 (Rev. ed.). Wellington, New Zealand: English Language Institute, Victoria University of Wellington.

Nation, I.S.P. (1990). *Teaching and learning vocabulary.* Boston, MA: Newbury House.

Nation, I.S.P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Nelson, K. (1977). The syntagmatic-paradigmatic shift revisited: A review of research and theory. *Psychological Bulletin, 84*(1), 93-116.

Newstead, S. E., Pollard, P., & Riezebos, D. (1987). The effects of set size on the interpretation of quantifiers used in rating scales. *Applied Ergonomics, 18,* 178-182.

Nissen, H. B., & Henriksen, B. (2006). Word class influence on word association test results. *International Journal of Applied Linguistics,16*(3), 389-408.

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning.* Cambridge: Cambridge University Press.

Odlin, T. (1993). Rediscovering interlanguage [Review of the book *Rediscovering interlanguage*]. *Language, 69*(2), 379-383.

Odlin, T. (2005). Crosslinguistic influence and conceptual transfer: What are the concepts? *Annual Review of Applied Linguistics, 25,* 3-25.

Odlin, T. (2008). Conceptual transfer and meaning extensions. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 306-340). New York: Routledge.

Orita, M. (2000). *Word Associations of Japanese EFL Learners and Native Speakers: Shift in Response Type Distribution and its Relevance to Lexical Development.* Unpublished MA thesis, the University of Birmingham, Birmingham, UK.

Orita, M. (2002a). Proficiency, lexical development, and the mental lexicon: Investigating the response type distribution of word associations of Japanese EFL learners and native speakers. *Yatsushiro Koosen Kiyoo, 24,* 113-124.

Orita, M. (2002b). Word associations of Japanese EFL learners and native speakers: Shifts in response type distribution and the associative development of individual words. *Annual Review of English Language Education in Japan, 13,* 131-120.

Orita, M. (2003). Atama no naka de tango wa donoyooni musubitsuite iru ka: Mentaru rekishikon no shikumi [How words are linked with each other in the brain: The organisation of the mental lexicon]. *Eigo Kyooiku, 52*(7), 14-16.

Osaka, S. (2005). The effects of categorical teaching of loanwords on vocabulary acquisition. *Annual Review of English Language Education in Japan, 16,* 161-170.

Oxford Advanced Learners' Dictionary (7th ed.). (2005). Oxford: Oxford University Press.

Pathfinder (Version 5.4) [Computer software and manual]. Retrieved 11 July, 2006, from http://www.interlinkinc.net/software/

Pease, D. M., Gleason, J. B., & Pan, B. A. (1993). Learning the meaning of words: Semantic development and beyond. In J. B. Gleason (Ed.), *The development of language* (pp. 115-149). New York: Macmillan.

Pollio, H. R. (1963). A simple matrix analysis of associative structure. *Journal of Verbal Learning and Verbal Behavior, 2*, 166-169.

Postman, L., & Keppel, G. (1970). *Norms of word associations*. New York: Academic Press.

Potter, M. C., So, K-F., von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23,* 23-38.

Preece, P. (1976). Mapping cognitive structure: A comparison of methods. *Journal of Educational Psychology, 68*(1), 1-8.

Preece, P. (1978). Exploration of semantic space: Review of research on the organization of scientific concepts in semantic memory. *Science Education, 62*(4), 547-562.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language.* London: Longman.

Racine, J. P. (2008). Cognitive processes in second language word association. *JALT Journal, 30*(1), 5-26.

Rapoport, A., & Fillenbaum, S. (1972). An experimental study of semantic structures. In A. K. Romney, R. N. Shepard & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences* (pp. 93-131). New York: Seminar Press.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10,* 355-371.

Read, J. (2000). *Assessing vocabulary.* Cambridge: Cambridge University Press.

Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 209-227). Amsterdam: John Benjamin.

Richard, J. C., & Schmidt, R. (2002). *Longman dictionary of language teaching and applied linguistics* (3rd ed.). London: Longman.

Riegel, K.F., Ramsey, R.M., & Riegel, R. M. (1967). A comparison of the first and second languages of American and Spanish students. *Journal of Verbal Learning and Verbal Behavior, 6,* 536-544.

Riegel, K.F., & Zivian, I.W.M. (1972). Study of inter- and intralingual associations in English and German. *Language Learning, 22,* 51-63.

Ringbom, H. (1983). On the distinctions of item learning vs. system learning and receptive competence vs. productive competence in relation to the role of L1 in foreign language learning. In H. Ringbom (Ed.), *Psycholinguistics and foreign language learning* (pp. 163-173), the Research Institute of the Åbo Akademi Foundation, Vol. 86.

Romney, A. K., Brewer, D. D., & Batchelder, W. H. (1993). Predicting clustering from semantic structure. *Psychological Science, 4*(1), 28-34.

Ronald, J. M. (2006). *Second Language Vocabulary Acquisition through Dictionary Use.* Unpublished PhD thesis, University of Wales Swansea, Swansea, UK.

Rosenberg, S. (1982). The method of sorting in multivariate research with applications selected from cognitive psychology and person perception. In N. Hirschberg & L. G. Humphreys (Eds.), *Multivariate applications in the social sciences* (pp. 117-142). Hillsdale, NJ: Lawrence Erlbaum.

Rosenberg, S., & Kim, M. P. (1975). The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research, 10*(4), 489-502.

Routh, D. A. (1994). On representations of quantifiers. *Journal of Semantics, 11,* 199-214.

Saito, T., & Yadohisa, H. (2006). *Kanrensei deeta no kaisekihoo: Tajigen shakudo kooseihoo to kurasutaa bunsekihoo* [Analysis of relational data: Multidimensional scaling and cluster analysis]. Tokyo: Kyoritsu Shuppan.

Sánchez, M. J. (2002). Efecto de la instrucción con patrones lingüísticos en el aprendizaje léxico: Campo semántico shine. *Estudios Ingleses de la Universidad Complutense, 10,* 183-199.

Sánchez, M. J. (2004). Effect of instruction with expert patterns on the lexical learning of English as a foreign language. *System, 32,* 89-102.

Sánchez-Casas, R., & García-Albea, J. (2005). The representation of cognate and noncognate words in bilingual memory: Can cognate status be characterized as a special kind of morphological relation? In J. F. Kroll & A.M.B. de Groot (Eds.) *Handbook of bilingualism: Psycholinguistic approaches* (pp. 226-250). New York: Oxford University Press.

Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika, 42*(3), 319-345.

Schmitt, N. (2000). *Vocabulary in language teaching.* Cambridge: Cambridge University Press.

Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in Second Language Acquisition, 19,* 17-36.

Schur, E. (2003). *An Exploration of the Structural Properties of L2 Vocabulary Networks: A Graph Theoretical Approach.* Unpublished PhD thesis, University of Wales Swansea, Swansea, UK

Schur, E. (2007). Insights into the structure of L1 and L2 vocabulary networks: Intimations of small worlds. In H. Daller, J. Milton & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 182-203), Cambridge: Cambridge University Press.

Schvaneveldt, R. W. (Ed.). (1990). *Pathfinder associative networks: Studies in knowledge organization.* Norwood, NJ: Ablex Publishing Corporation.

Selinker, L. (1972). Interlanguage. *IRAL, 10*(2), 209-231.

Selinker, L., & Lakshamanan, U. (1992). Language transfer and fossilization: The multiple effects principle. In S. M. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 197-216). Amsterdam: John Benjamins.

Singleton, D. (1999). *Exploring the second language mental lexicon.* Cambridge: Cambridge University Press.

Singleton, D. (2000). *Language and the lexicon: An introduction.* London: Arnold.

Skehan, P. (1991). Individual differences in second language learning. *Studies in Second Language Acquisition, 13*, 275-298.

Smith, M. C. (1997). How do bilinguals access lexical information? In A.M.B. de Groot & J.F. Kroll (Eds.), *Tutorials in bilingualism* (pp.145-168). Mahwah, NJ: Lawrence Erlbaum.

Sonoda, K. (1996). *Hokkaido University vocabulary list.* Sapporo: The Institute of Language and Culture Studies, Hokkaido University.

Steinberg, D. D. (1967). The word sort: An instrument for semantic analysis. *Psychometric Science, 12,* 541-542.

Stevenson, R. (1883). *Treasure Island*, "Chapter 32. The Treasure-hunt : The Voice Among the Trees", Electronic Text Center, University of Virginia Library. Retrieved 4 December 2005, from http://etext.lib.virginia.edu/toc/modeng/public/SteTrea.html

Sunderman, G., & Kroll, J. F. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquisition, 28,* 387-422.

Szalay, L. B., & Deese, J. (1978). *Subjective meaning and culture: An assessment through word associations.* Hillsdale, NJ: Lawrence Erlbaum.

Söderman, T. (1989). Word associations of foreign language learners and native speakers: A shift in response type and its relevance for a theory of lexical development. *Scandinavian Working Papers on Bilingualism, 8,* 114-121.

Söderman, T. (1993). Word associations of foreign language learners and native speakers: The phenomenon of a shift in response type and its relevance for lexical development. In H. Ringbom (Ed.), *Near-native proficiency in English* (pp. 91-182). *English Department Publications 2,* Åbo Akademi University, Åbo, Finland.

Sökmen, A. J. (1993). Word association results: A window to the lexicons of ESL students.

JALT Journal, 15, 135-150.

Sökmen, A. J. (1997). Current trends in teaching second language vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 237-257), Cambridge: Cambridge University Press.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson.

Takane, Y. (1980). Analysis of categorizing behavior by a quantification method. *Behaviormetrika. 8,* 75 - 86.

Takemura, A. (1991). *Gendai Suuri Tookeigaku* [Modern Mathematical Statistics]. Tokyo: Sobunsha.

Takeuchi, H., & Utsugi, A. (1988). An analysis of pain concepts by network model. *Mathematical Linguistics, 16,* 233-245.

Tamamura, F. (1991). Nihongo ni okeru gairaiyooso to gairaigo [Usage of foreign language elements and loan words in the Japanese language]. *Nihongo Kyooiku, 74,* 13-27.

Tanaka, S., & Yamagiwa, Y. (1989). *Shinpan yuuzaa no tame no kyooiku/shinritookei to jikkenkeikakuhoo* [Educational and psychological statistics and research design (New ed.)]. Tokyo: Kyoiku Shuppan.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327-352.

Waern, Y. (1972). Structure in similarity matrices. *Scandinavian Journal of Psychology, 13,* 5-16.

Watts, D. J. (1999). *Small world: The dynamics of networks between order and randomness.* Princeton, NJ: Princeton University Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393,* 440-442.

White, H. D. (2003). Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists. *Journal of the American Society for Information Science and Technology, 54(5),* 423-434.

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research, 12*(3), 233-265.

Wilks, C. (1999). *Untangling Word Webs: Graph Theory Approaches to L2 Lexicons.* Unpublished PhD thesis, University of Wales Swansea, Swansea, UK.

Wilks, C., & Meara, P. (2002). Untangling word webs: Graph theory and the notion of density in second language word association networks. *Second Language Research, 18*(4), 303-324.

Wilks, C., & Meara, P. (2007). Implementing graph theory approaches to the exploration of density and structure in L1 and L2 word association networks. In D. Helmut, J. Milton & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp.

167-181), Cambridge: Cambridge University Press.

Wilks, C., Meara, P., & Wolter, B. (2005). A further note on simulating word association behaviour in a second language. *Second Language Research, 21*(4), 359-372.

Wolter, B. (2001). Comparing the L1 and the L2 mental lexicon. *Studies in Second Language Acquisition, 23,* 41-69.

Wolter, B. (2002). Assessing proficiency through word associations: Is there still hope? *System, 30,* 315-329.

Wolter, B. (2005). *V_Links: A New Approach to Assessing Depth of Word Knowledge.* Unpublished PhD thesis, University of Wales Swansea, Swansea, UK.

Wolter, B. (2006). Lexical network structures and L2 vocabulary acquisition: The role of L1 lexical/conceptual knowledge. *Applied Linguistics 27*(4), 741-747.

Wray, A., Trott, K., & Bloomer, A. (1998). *Projects in linguistics.* London: Arnold.

Yamashita, J. (2007). Investigating asymmetry in EFL learners' mental lexicon: Connections between lexical and conceptual representations in L1 and L2. *JACET Journal, 45,* 63-79.

Yokokawa, H., Yabuuchi, S., Kadota, S., Nakanishi, Y., & Noro, T. (2002). Lexical networks in L2 mental lexicon: Evidence from a word-association task for Japanese EFL learners. *Language Education & Technology, 39,* 21-39.

Yu, L. (1996). The role of cross-linguistic lexical similarity in the use of motion verbs in English by Chinese and Japanese learners. Unpublished Ed. D. dissertation, University of Toronto, Canada.

Appendices

Appendix 3.1. Task 1 (Primary Sorting Task)

*Instructions**: There are 39 English adjectives below. They can be categorised into one of the four groups of adjectives: (a) WEIGHT and SIZE, (b) TEMPERATURE, (c) HAPPINESS and (d) ANXIETY. Sort the 39 adjectives into the four groups by writing each adjective into the column you think it best fits into. Put all adjectives into one of the groups. Note that the number of adjectives in each of the four groups will likely not be the same.

39 adjectives:

afraid, anorexic, anxious, cheerful, chilled, chubby, cold, excited, fat, flabby, freezing, frightened, gay, glad, happy, hot, jolly, lean, lukewarm, nervous, pleased, plump, podgy, portly, scalding, scared, scrawny, skinny, slender, slim, spare, stout, tepid, thin, tubby, uneasy, warm, wiry, worried.

① WEIGHT and SIZE adjectives	② TEMPERATURE adjectives	③ HAPPINESS adjectives	④ ANXIETY adjectives

*The instructions and the subset names were written in Japanese. This was also the case with Tasks 2 and 3.

Appendix 3.2. Task 2 (Card-Sorting Task)

Instructions: There are 30 adjectives of emotion below. They can be sorted into four groups. First, put the adjectives into one of the columns of either 1①, 1②, 1③ or 1④. Put the adjectives that you don't know into the 1⑤ column. Note that the number of adjectives in each of the four groups will likely not be the same. Then, write the name of the group of adjectives into columns 2①, 2②, 2③, 2④, respectively, in Japanese.

30 adjectives of emotion:

alarmed, annoyed, anxious, cheerful, chuffed, cross, depressed, disappointed, distressed, elated, excited, frightened, furious, glad, grumpy, high, mad, miserable, moody, outraged, overjoyed, panic-stricken, petrified, pleased, scared, sorrowful, terrified, thrilled, uneasy, upset.

	1①	1②	1③	1④	1⑤ adjectives you don't know
adjective					
Name of the group of adjectives	2①	2②	2③	2④	

Appendix 3.3. Task 3 (Situation Task)
Instructions: Below are 16 situations described in English. Circle the adjective(s) that you feel have more or less the same meaning as the underlined one. Note that in some cases you are expected to circle one word, and in others, two or three.

1. Your parents decide to give you money to buy new furniture for your room.
   You are/feel <u>happy</u>.

   chuffed
   excited
   pleased
   haunted

2. Your dog is run over by a car and dies.
   You are/feel <u>sad</u>.

   distressed
   sorrowful
   content
   miserable

3. Your brother arrives an hour late so you miss going to an exciting rock concert.
   You are/feel <u>angry</u>.

   impressed
   flabbergasted
   mad
   annoyed

4. You have tried to get on the best sports team in town for more than two years. Today you get a letter saying that you can start on the team tomorrow.
   You are/feel <u>happy</u>.

   grieved
   desperate
   high
   thrilled

5. You are alone in the house one night and suddenly you hear footsteps in the hall.
   You are/feel <u>afraid</u>.

   startled
   panic-stricken
   suspicious
   worried

6. Your best friend chooses to go on holiday with another friend instead of with you.
   You are/feel <u>sad</u>.

   delighted
   disappointed
   moody

left out

7. You are out skiing and suddenly get caught in an avalanche.
You are/feel afraid.

alarmed
anxious
shocked
scared

8. Your family is moving to another part of the country and you have to leave all your friends.
You are/feel sad.

inquisitive
puzzled
lonely
depressed

9. You are on the way to the movies with a group of friends.
You are/feel happy.

cheerful
gay
betrayed
thoughtful

10. Your mother has promised you not to tell anybody about your new girl/boyfriend. And then you hear her tell the neighbour all about it!
You are/feel angry.

upset
relieved
proud
outraged

11. You are staying in an old wooden farmhouse. You wake up in the middle of the night. The house is on fire!!
You are/feel afraid.

terrified
petrified
frightened
rejected

12. You have been waiting in a queue for some tickets for three hours. A girl suddenly pushes her way to the front of the queue and buys the last two tickets that you should have had.
You are/feel angry.

astonished
jealous

helpless
furious

13. Your parents won't let you go to a party with your friends because you haven't cleaned your room.
You are/feel <u>angry</u>.

cross
grumpy
radiant
ecstatic

14. You inherit £ 1 million from an old aunt.
You are/feel <u>happy</u>.

glad
overjoyed
heartbroken
elated

15. You have had a toothache for a week. You are now going to the dentist.
You are/feel <u>afraid</u>.

shaky
distraught
nervous
uneasy

16. You hear on television that your favourite pop group plans to stop.
You are/feel <u>sad</u>.

indifferent
upset
frustrated
dazzle

Appendix 4.1. Directions for Card-sorting Game

In the envelope there are 50 cards with different English words printed on them. You will sort all of them into groups of words that go together according to meaning. You can decide the categories yourself. (Don't sort words by parts of speech (e.g., nouns, verbs, etc.) There might be a few words that don't seem to fit in any of your groups. You can leave these as single cards. It doesn't matter how many groups you make.

Use your imagination and fit each word into the groups you have created. Remember, this is a game. Just enjoy doing it!

You will have 20 minutes. You should have time to think it over and change it at the end before you stop. You will get a small present when you finish.

Appendix 4.2a. Co-occurrence (Frequency count) matrix for NS (n = 28) (k = 50)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	*ago*	28		5	19	8	1	3	3	2	2		16			3	2	1		17	3	1	2	2	10	2	2	4	1	1	15	3		3	1	1		3	1	1	1	4	2		1	18	1	4	1	4	14
2	*air*		28	4	2		7	2	1		1	3		11		4			3	1	2	3	8		1	2	1	2	9	21	1	3	5	3		5		2	1	1	2		2	3	3	1		1	3	1	
3	*all*	5	4	28	5	7	2		3	4	2	1		3		1		1	4	2	3	2	6	2	2	1	2	11	6	2	5	11		3	2	1		2	2	1	3	1	2	1	1	5	4	12	2		6
4	*already*	19	2	5	28	12	2	1	4		1		8	2		2	1	1		14	1	1	4	1	4	3	1	3	2	4	16	2		4	2	2			1	2	1	2	2		1	18	1	5	3	1	18
5	*also*	8		7	12	28	2	2	2	1			1	1	1	1			1	4	2		4	2	1	2		5	4	1	11	2		4	3	2	1			2	1	1	1		1	12	2	7	2	1	15
6	*area*	1	7	2	2	2	28	1	6			9		6	7	8			1	4	2	3	5	1		3	1	10	5	7	3	2	8	1	2	22		1		1	17	1	2	3	19	1	1	1	2	2	3
7	*arm*	3	2		1	2	1	28		2	7	2	5	1	3	2	6	6		1	4		2	1	6	3	9		4	1	1	2		3	4		6	5	5	1		3	3	3	1	2	2	2	2	7	1
8	*arrive*	3	1	3	4	2	6		28	3		3	2	5	7	1	2		3	10	3	9	2			4		1	2	1	5	2	6	1	3	3				4	4		9	8	4	2	4	1	11	1	3
9	*believe*	2		4		1		2	3	28	4		1	8	3	2	12	1	18		8	10	3	10	2	5	2	1	4	1		4	3	4	3		1	5	2	16		5	3	4			22	1	3	2	
10	*boy*	2	1	2	1			7		4	28			1			7	12	2	1	7	1		4		2		1		2	1	1		6	19		8		7			5	2	1		1	2	1	3	1	
11	*business*		3	1			9	2	3			28	5	1	8	8		3	2	3	1	1	4	1	8	1	8	7	2	2		1	8	1		10	3	7	6		15	7	2	1	10			1		8	
12	*century*	16			8	1		5	2	1		5	28		1	9	3	4		11	2		1		17	1	7	1			9	3	1	1	1	1	2	6	6	2	1	4	2	1	2	9	1	1	1	12	5
13	*clear*		11	3	2	1	6	1	5	8	1	1		28	4	2	3		8		2	8	3	4		4			6	10	2	4	8	2	1	3		2		11	1	2	1	5	2	1	10	1	3	1	2
14	*close*					1	7	3	7	3		8	1	4	28	1	3	1	3	5		6	1	3		6	1	5	1	1	2	3	19	2		6	2	2		3	9	3	3	5	8	1	3	3	1		1
15	*country*	3	4	1	2	1	8	2	1	2		8	9	2	1	28		3	1	2	2	1	3	2	12	2	10	6	3	6	4		4	3	1	9	5	10	7	5	7	7	1		7	1	3		3	11	1
16	*cry*	2			1			6	2	12	7		3	3	3		28	4	8	2	3	4		8	2	5	1		2	2	1	2	2	3	4		3	4	2	6		9	6	6	1	3	9	1	6	3	
17	*doctor*	1		1	1			6		1	12	3	4		1	3	4	28			3		1	7	4		8		2	1	1	2	1	3	12		18	3	14			7	2	1	1	1	2	1	1	3	
18	*dream*		3	4		1	1		3	18	2	2		8	3	1	8		28		4	6	3	7	2	2	1		6	3		4	3	3	2			5	2	11	1	3	1	1		1	16	1	1		1
19	*early*	17	1	2	14	4	4	1	10		1	3	11		5	2	2			28	2	2	1	1	4	2		3	1	2	14	2	6	1	1	3		2			4	3	3	1	3	12		2	2	3	9
20	*figure*	3	2	3	1	2	2	4	3	8	7	1	2	2		2	3	3	4	2	28	6	7	4	2	2	1	3	7	3	1	2	1	4	10	4	2	2	2	6	2	3	3	1	1	1	7	1	2	3	1
21	*find*	1	3	2	1		3		9	10	1	1		8	6	1	4		6	2	6	28	3	5		11		4	2	3		3	6	2		4		1		8	2	2	3	4	3	1	9		3		
22	*form*	2	8	6	4	4	5	2	2	3		4	1	3	1	3		1	3	1	7	3	28	3	1	2	3	3	12	8	2	6	2	4	3	5	1	3	1	6	2	3	5	1	2	3	5	3	2	4	3
23	*help*	2		2	1	2	1	1		10	4	1		4	3	2	8	7	7	1	4	5	3	28	2	8	6	2	5	2	2		2	7	5	1	5	3	2	7	1	11	4	3	2	1	10	2	2	1	1
24	*history*	10	1	2	4	1		6		2		8	17			12	2	4	2	4	2		1	2	28	1	12	1	1		2	1		2	1	1	5	12	10	2	2	8	1		1	5	2		1	17	3
25	*keep*	2	2	1	3	2	3	3	4	5	2	1	1	4	6	2	5		2	2	2	11	2	8	1	28	3	4	2	2	3	1	3	6	1	3	3	2	1	4	3	4	5	4	4	2	3	3	6	1	2
26	*law*	2	1	2	1		1	9		2		8	7		1	10	1	8	1		1		3	6	12	3	28		3	1		2		1			15	11	10	3		9	4	1	2	1	2	3	1	12	
27	*lot*	4	2	11	3	5	10		1	1	1	7	1		5	6				3	3	4	3	2	1	4		28	1	2	2	4	5	4	1	12					11	1	1		9	3		8			3
28	*matter*	1	9	6	2	4	5	4	2	4		2		6	1	3	2	2	6	1	7	2	12	5	1	2	3	1	28	9	2	6	2	2	3	4	1	2	1	5	1	3	1	1	1	2	6	2	1	1	3
29	*nature*	1	21	2	4	1	7	1	1	1	2	2		10	1	6	2	1	3	2	3	3	8	2		2	1	2	9	28	2	2	7	3	1	6		4		4	1	2	3	1	2	2	3	1	3	2	1
30	*next*	15	1	5	16	11	3	1	5		1		9	2	2	4	1	1		14	1		2	2	2	3		2	2	2	28	1	1	4	3	2	2	2	2	2	2	3	1	1	1	16	1	4	3	3	16
31	*nothing*	3	3	11	2	2	2	2	2	4	1	1	3	4	3		2	2	4	2	2	3	6		1	1	2	4	6	2	1	28	3	1		3	2	2		3	1	1	1	2	3	6	4	7	1	3	3
32	*open*		5				8		6	3		8	1	8	19	4	2	1	3	6	1	6	2	2		3		5	2	7	1	3	28			7		1		3	8	1	1	3	7		3		1	1	
33	*our*	3	3	3	4	4	1	3	1	4	6	1	1	2	2	3	3	3	3	1	4	2	4	7	2	6	1	4	2	3	4	1		28	7	2	4	1	3	3	2	4	3		2	3	2	3	3	2	3
34	*person*	1		2	2	3	2	4	3	3	19		1	1		1	4	12	2	1	10		3	5	1	1		1	3	1	3			7	28	2	9	2	9	3	1	5	5	2	1	1	3	1	5	3	2
35	*place*	1	5	1	2	2	22		3			10	1	3	6	9				3	4	4	5	1	1	3		12	4	6	2	3	7	2	2	28		1		2	19	2	1	3	21	1	1	1	2	1	2
36	*police*					1		6		1	8	3	2		2	5	3	18			2		1	5	5	3	15		1		2	2		4	9		28	7	17	1		8	2	1	3	2	1	3	1	6	1
37	*power*	3	2	2			1	5		5		7	6	2	2	10	4	3	5	2	2	1	3	3	12	2	11		2	4	2	2	1	1	2	1	7	28	14	7	1	10	2	1		1	4	1		17	
38	*president*	1	1	2	1			5		2	7	6	6			7	2	14	2		2		1	2	10	1	10		1		2			3	9		17	14	28	2	1	5	1				2		1	10	
39	*reason*	1	1	1	2	2	1	1	4	16			2	11	3	5	6		11		6	8	6	7	2	4	3		5	4	2	3	3	3	3	2	1	7	2	28		5	2	3		1	20		2	5	1
40	*shop*	1	2	3	1	1	17		4			15	1	1	9	7			1	4	2	2	2	1	2	3		11	1	1	2	1	8	2	1	19		1	1		28	2	1	2	19	1		1			2
41	*social*	4		1	2	1	1	3		5	5	7	4	2	3	7	9	7	3	3	3	2	3	11	8	4	9	1	3	2	3	1	1	4	5	2	8	10	5	5	2	28	6	4	3	3	4	3	1	9	1
42	*stand*	2	2	2	2	1	2	3	9	3	2	2	2	1	3	1	6	2	1	3	3	3	5	4	1	5	4	1	1	3	1	1	1	3	5	1	2	2	1	2	1	6	28	18	3	3	3	4	20	1	1
43	*step*		3	1			3	3	8	4	1	1	1	5	5		6	1	1	1	1	4	1	3		4	1		1	1	1	2	3		2	3	1	1		3	2	4	18	28	5	2	3	3	20		
44	*street*	1	3	1	1	1	19	1	4			10	2	2	8	7	1	1		3	1	3	2	2	1	4	2	9	1	2	1	3	7	2	1	21	3				19	3	3	5	28	3		3	2	1	1
45	*then*	18	1	5	18	12	1	2	2		1		9	1	1	1	3	1	1	12	1	1	3	1	5	2	1	3	2	2	16	6		3	1	1	2	1		1	1	3	3	2	3	28		6	1	3	17
46	*understand*	1		4	1	2	1	2	4	22	2		1	10	3	3	9	2	16		7	9	5	10	2	3	2		6	3	1	4	3	2	3	1	1	4	2	20		4	3	3			28	1	3	2	1
47	*very*	4	1	12	5	7	1	2	1	1	1	1	1	1	3		1	1	1	2	1		3	2		3	3	8	2	1	4	7		3	1	1	3	1			1	3	4	3	3	6	1	28	2		6
48	*walk*	1	3	2	3	2	2	2	11	3	3		1	3	1	3	6	1	1	2	2	3	2	2	1	6	1		1	3	3	1	1	3	5	2	1		1	2		1	20	20	2	1	3	2	28	1	2
49	*war*	4	1		1	1	2	7	1	2	1	8	12	1		11	3	3		3	3		4	1	17	1	12		1	2	3	3	1	2	3	1	6	17	10	5		9	1		1	3	2		1	28	1
50	*while*	14		6	18	15	3	1	3				5	2	1	1			1	9	1		3	1	3	2		3	3	1	16	3		3	2	2	1			1	2	1	1		1	17	1	6	2	1	28

Note. Blank cells equal "0".

Appendix 4.2b. Co-occurrence (Frequency count) matrix for NNS (n = 28) (k = 50)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	*ago*	28	2	1	21	1	2	3	2	2	1	2	13		3	4	2	1	1	21	3	1	1	3	11	1	3		3	2	19	3	1	3		1	2	4	4	5	1	3				22	1	1		5	20
2	*air*	2	28	3	1		6	2			2		1	11	2	9	2	2	1		3		4	3	3	3	1		3	20		4	3	2	2	6	2	2	2	3	4	2	1	2	5	1	1		2	2	
3	*all*	1	3	28	3	10		2	2	2	1	1	2	2		2	2	1	2	2	4	2	5	3	1	3	2	16	2	2	2	16		5			2	2	2	3		3	2	1		3	3	13		2	1
4	*already*	21	1	3	28	4		1	3			1	10	1	3	1	1			23	3	1	1	1	7	1	1	1	2	1	18	3	1	1			1	1	1	3	1	1				23	2	2		1	21
5	*also*	1		10	4	28		1	1			1				1	1	1		3				1			1	8			4	7		2			1	1	1	1		1		1		5	2	9		1	3
6	*area*	2	6				28	2	6			3	2	1	4	10	1	1	1			2	3	1	4	2	2	6	4	5	2	1	4	2	1	23	2	4	3	4	14	2	5	9	21				6	4	
7	*arm*	3	2	2	1	1	2	28	2	4	6	3	2	2	3	4	4	7	4		10	2	7	8	6	4	8	1	3	2		1	3	7	7		7	9	9	4	1	6	4	1	1	1	5		3	9	
8	*arrive*	2		2	3	1	6	2	28	6	1	4	2	4	11	1	6		4	4	3	9	4	6	1	7		3	1		6	1	10	1	1	6				1	7		16	8	9	4	6	2	17		3
9	*believe*	2		2				4	6	28	7	1	4	6	3	2	10	3	17	1	6	15	2	8	3	9	1	1	2		2		4	3	3	1		3	2	4		1	5	2		1	14	2	4	3	
10	*boy*	1	2	1				6	1	7	28	4	3	2	1	2	4	15	9	1	6	3	1	1	2		2			2	3			6	17	1	9	2	11	1	4	3	1	4	2	1	1	1		3	
11	*business*	2		1	1	1	3	3	4	1	4	28	2		4	5		8	5	2	4	1	2	1	5	3	12	3	2	1	2		6	5	7	3	9	8	10	1	11	11	2	4	6	1	1		4	10	1
12	*century*	13	1	2	10		2	2	2	4	3	2	28		1	5	2		5	10	4	3	2	2	18	2	5		4	2	10	2		4	1	1	2	8	5	5		5	1	2	2	11	2	1		9	10
13	*clear*		11	2	1		1	2	4	6	2			28	8	5	6	2	3	2	5	8	3	6		7	1	1	4	11		1	9	1	3	2	2	1	1	1	1	2	5	1	1		8		6		
14	*close*	3	2		3		4	3	11	3	1	4	1	8	28	2	5		5	4	4	6	5	4		6	2	2	2	2	6		22		1	4	1	1	1		7	1	11	7	6	3	4		14		3
15	*country*	4	9	2	1	1	10	4	1	2	2	5	5	5	2	28	3	3	5		1		1	3	9		9	3	2	10		2	3	5	2	9	8	10	8	4	5	11	2	2	7	1	2		2	12	
16	*cry*	2	2	2	1	1	1	4	6	10	4		2	6	5	3	28	4	8	2	3	6	3	15	5	6	2		5	2	1	5	6		1	1	1	4	1	7		1	8	1		1	6	2	8	7	1
17	*doctor*	1	2	1		1	1	7		3	15	8		2		3	4	28	3	2	2	1	1	6	2	1	5	1	2	2		1		6	15	1	16	4	15	2	4	6			2		2			6	
18	*dream*	1	1	2			1	4	4	17	9	5	5	3	5	5	8	3	28		7	11	3	7	6	7	3	3	2	1	4		4	3	3	3	1	5	3	2	2	3	6	5	2	1	7	2	5	6	
19	*early*	21		2	23	3			4	1	1	2	10	2	4		2	2		28	2	1		1	6	1		1	1		19	2	2		1					2	2	1			1	22	1	2	1		21
20	*figure*	3	3	4	3			10	3	6	6	4	4	5	4	1	3	2	7	2	28	8	13	5	2	5	4	1	6	4	3	3	4	4	4		2	3	2	7	1	3	5	5		4	6	1	2	3	2
21	*find*	1		2	1		2	2	9	15	3	1	3	8	6		6	1	11	1	8	28	3	9	1	11		3	4		4	1	6	1	1	3	1			4	1		6	5	2	2	17	1	6		1
22	*form*	1	4	5	1		3	7	4	2	1	2	2	3	5	1	3	1	3		13	3	28	5	2	8	6	1	10	4	1	3	4	3	3	1	4	5	4	6	1	6	7	6	2	1	5		5	4	
23	*help*	3	3	3	1	1	1	8	6	8	1	1	2	6	4	3	15	6	7	1	5	9	5	28	4	8	4	1	6	3		4	5	2		1	4	4	3	5		3	7			1	9		6	5	
24	*history*	11	3	1	7		4	6	1	3	2	5	18			9	5	2	6	6	2	1	2	4	28	2	10		5	4	5	3		4	1	1	5	14	8	7	1	9	1		2	7	2			17	6
25	*keep*	1	3	3	1		2	4	7	9		3	2	7	6		6	1	7	1	5	11	8	8	2	28	4	3	6	2	1	3	7	1	1	1	2	2	2	4	2	2	8	2	3	1	7		7	2	1
26	*law*	3	1	2	1	1	2	8		1	2	12	5	1	2	9	2	5	3		4		6	4	10	4	28	1	7	2		1	2	5	4		17	18	13	8	3	15	1	2	2	1	3		1	17	
27	*lot*			16	1	8	6	1	3	1		3		1	2	3		1	3	1	1	3	1	1		3	1	28			1	9	1	2		6	1	1	1		6	1	2	2	7		1	11	2	1	
28	*matter*	3	3	2	2		4	3	1	2		2	4	4	2	2	5	2	2	1	6	4	10	6	5	6	7		28	4		5	2	3	1	1	7	6	4	15		6	1	3	1	2	7		2	5	1
29	*nature*	2	20	2	1		5	2			2	1	2	11	2	10	2	2	1		4		4	3	4	2	2		4	28		3	2	3	2	5	2	4	2	4	1	3	1	1	2	1	1		2	3	
30	*next*	19		2	18	4	2		6	2	3	2	10		6		1		4	19	3	4	1		5	1		1			28	1	3	1	1	3				1	3		3	6	2	23	1	4	2		17
31	*nothing*	3	4	16	3	7	1	1	1				2	1		2	5	1		2	3	1	3	4	3	3	1	9	5	3	1	28		2		1	1	1	1	6		1	1			3	2	7		3	2
32	*open*	1	3		1		4	3	10	4		6		9	22	3	6		4	2	4	6	4	5		7	2	1	2	2	3		28		1	4	1	1	2		10	2	11	5	6	1	5		14		1
33	*our*	3	2	5	1	2	2	7	1	3	6	5	4	1		5		6	3		4	1	3	2	4	1	5	2	3	3	1	2		28	5	1	6	6	8	2	1	12	1	1	1	2	3	2		6	
34	*person*		2				1	7	1	3	17	7	1	3	1	2	1	15	3	1	4	1	3		1	1	4		1	2	1		1	5	28	3	12	3	13	1	4	5	1	2	3		2		1	3	
35	*place*	1	6				23		6	1	1	3	1	2	4	9	1	1	3			3	1	1	1	1		6	1	5	3	1	4	1	3	28		1		2	14		5	11	20				7	1	
36	*police*	2	2	2	1	1	2	7			9	9	2	2	1	8	1	16	1		2	1	4	4	5	2	17	1	7	2		1	1	6	12		28	13	18	5	5	13			3	1	3			12	
37	*power*	4	2	2	1	1	4	9		3	2	8	8	1	1	10	4	4	5		3		5	4	14	2	18	1	6	4		1	1	6	3	1	13	28	14	9	1	13	1	1	1	1	3			19	
38	*president*	4	2	2	1	1	3	9		2	11	10	5	1	1	8	1	15	3		2		4	3	8	2	13	1	4	2		1	2	8	13		18	14	28	4	7	12			3	1	3			13	
39	*reason*	5	3	3	3	1	4	4	1	4	1	1	5	1		4	7	2	2	2	7	4	6	5	7	4	8		15	4	1	6		2	1	2	5	9	4	28		5		4	1	3	6	1	1	8	2
40	*shop*	1	4		1		14	1	7		4	11		1	7	5		4	2	2	1	1	1		1	2	3	6		1	3		10	1	4	14	5	1	7		28	4	5	7	19	1			7	2	1
41	*social*	3	2	3	1	1	2	6		1	3	11	5	2	1	11	1	6	3	1	3		6	3	9	2	15	1	6	3		1	2	12	5		13	13	12	5	4	28		1	2	1	4			13	
42	*stand*		1	2			5	4	16	5	1	2	1	5	11	2	8		6		5	6	7	7	1	8	1	2	1	1	3	1	11	1	1	5		1			5		28	12	6	1	3	1	19	1	
43	*step*		2	1		1	9	1	8	2	4	4	2	1	7	2	1		5		5	5	6			2	2	2	3	1	6		5	1	2	11		1		4	7	1	12	28	10	2	1	1	12	1	
44	*street*		5				21	1	9		2	6	2	1	6	7		2	2	1		2	2		2	3	2	7	1	2	2		6	1	3	20	3	1	3	1	19	2	6	10	28				10	2	
45	*then*	22	1	3	23	5		1	4	1	1	1	11		3	1	1		1	22	4	2	1	1	7	1	1		2	1	23	3	1	2			1	1	1	3	1	1	1	2		28	2	5		1	21
46	*understand*	1	1	3	2	2		5	6	14	1	1	2	8	4	2	6	2	7	1	6	17	5	9	2	7	3	1	7	1	1	2	5	3	2		3	3	3	6		4	3	1		2	28	1	4	2	1
47	*very*	1		13	2	9			2	2	1		1				2		2	2	1	1						11			4	7		2						1			1	1		5	1	28			2
48	*walk*		2				6	3	17	4		4		6	14	2	8		5	1	2	6	5	6		7	1	2	2	2	2		14		1	7				1	7		19	12	10		4		28		
49	*war*	5	2	2	1	1	4	9		3	3	10	9			12	7	6	6		3		4	5	17	2	17	1	5	3		3		6	3	1	12	19	13	8	2	13	1	1	2	1	2			28	
50	*while*	20		1	21	3			3			1	10		3		1			21	2	1			6	1			1		17	2	1							2	1					21	1	2			28

Note. Blank cells equal "0".

Appendix 4.3a. Number of clustered words each word made for NS (k = 49)

	Participant																												Average
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	
1 *ago*	7	6	7	4	5	8	7	4	9	9	10	6	7	6	9	3	8	6	5	12	4	8	5	11	6	9	23	12	7.71
2 *air*	3	8	3	4	2	5	4	2	9	8	8	4	1	7	4	2	9	5	3	9	7	9	3	7	6	10	2	9	5.46
3 *all*	2	4	7	4	2	5	5	3	11	8	10	1	2	3	9	4	7	5	3	9	2	5	2	11	7	7	23	12	6.18
4 *already*	7	6	7	4	5	8	7	2	9	17	8	6	7	2	9	3	8	6	5	1	5	8	5	11	6	10	23	12	7.39
5 *also*	1	4	7	4	3	8	4	4	9	17	1	6	1	3	9	2	10	1	5	1	5	1	1	11	6	1	23	12	5.71
6 *area*	7	8	10	5	4	8	9	5	9	17	7	5	3	7	8	5	7	5	4	12	7	9	3	7	3	7	23	9	7.61
7 *arm*	4	1	3	1	1	3	4	4	9	14	10	1	3	5	4	3	8	7	7	12	11	1	12	4	5	4	5	9	5.54
8 *arrive*	2	4	5	3	5	7	7	1	9	17	10	3	8	6	8	2	7	5	4	12	1	7	3	5	6	7	10	12	6.29
9 *believe*	6	10	4	4	6	6	6	3	11	9	10	6	8	4	9	2	10	12	6	12	4	4	12	11	7	7	10	14	7.61
10 *boy*	5	5	3	3	3	3	5	2	9	9	5	1	3	3	9	3	4	6	5	10	2	5	12	4	6	7	5	6	5.11
11 *business*	7	5	10	6	2	8	10	5	7	8	6	3	5	5	8	5	7	7	2	12	11	9	4	7	6	7	10	9	6.82
12 *century*	7	5	7	4	5	8	7	4	7	14	10	4	7	6	7	5	8	7	5	12	4	8	5	3	7	9	10	9	6.93
13 *clear*	2	10	4	3	1	6	4	3	9	17	10	4	2	7	4	3	7	12	6	9	7	3	3	7	6	10	10	14	6.54
14 *close*	2	3	10	3	2	8	9	1	9	14	7	3	5	5	8	2	7	12	1	12	1	7	2	5	6	7	10	14	6.25
15 *country*	7	3	10	5	4	7	10	5	11	17	8	3	10	7	7	5	9	5	5	12	11	9	6	3	7	9	10	9	7.64
16 *cry*	6	10	1	4	6	3	6	1	9	14	10	4	8	4	9	1	4	12	1	10	1	1	12	11	5	7	5	14	6.39
17 *doctor*	5	5	3	4	3	7	5	3	7	14	5	5	6	4	3	2	8	6	5	10	11	5	12	3	5	2	10	6	5.86
18 *dream*	6	10	2	4	6	7	6	3	11	8	10	6	2	4	3	2	10	12	2	5	4	4	1	7	6	7	10	14	6.14
19 *early*	7	6	5	4	5	8	7	2	9	2	10	3	7	6	8	3	9	6	5	12	4	8	3	5	6	9	23	12	6.93
20 *figure*	6	8	7	3	2	7	4	4	1	9	10	6	6	3	9	3	9	6	6	12	2	2	4	4	7	7	23	6	6.29
21 *find*	2	3	2	3	1	2	9	3	9	9	10	1	8	6	3	2	9	12	6	12	2	3	2	11	7	10	10	14	6.11
22 *form*	3	8	10	4	1	5	10	4	7	17	6	6	1	6	7	3	10	5	2	9	2	2	1	11	7	10	23	9	6.75
23 *help*	2	10	4	4	6	7	6	3	11	9	6	4	6	5	9	2	10	12	5	10	2	3	1	11	6	2	23	14	6.89
24 *history*	4	5	7	6	2	7	10	4	11	8	10	4	10	6	7	5	8	7	5	12	11	8	5	3	6	9	10	9	7.11
25 *keep*	2	10	2	3	1	2	9	1	9	9	8	1	8	5	7	1	4	1	7	12	2	7	2	11	6	10	23	14	6.32
26 *law*	4	5	4	4	2	7	10	5	11	14	6	3	10	4	7	4	8	7	7	12	11	5	6	3	6	4	10	9	6.71
27 *lot*	7	4	10	4	2	8	9	3	1	9	1	5	5	3	9	4	9	5	3	12	1	9	2	11	3	1	23	9	6.14
28 *matter*	3	8	3	4	1	5	4	4	11	17	10	4	6	7	7	3	10	5	2	5	2	5	1	7	5	4	23	14	6.43
29 *nature*	3	8	3	4	2	5	4	2	9	17	8	3	6	7	7	2	9	12	3	9	7	9	3	7	6	10	2	9	6.29
30 *next*	7	6	7	4	3	8	7	4	9	17	8	6	7	2	9	2	7	6	5	10	4	8	3	5	7	9	23	12	7.32
31 *nothing*	2	8	2	4	1	6	9	3	7	14	10	4	2	1	7	4	5	5	1	5	2	5	1	11	7	7	10	12	5.54
32 *open*	2	3	10	3	2	8	4	3	7	2	7	3	5	7	8	2	9	12	3	12	7	7	2	5	6	7	10	14	6.07
33 *our*	2	1	1	1	1	7	5	1	9	9	8	1	10	5	9	1	10	1	1	10	1	1	12	11	6	10	23	9	5.93
34 *person*	5	5	7	3	3	7	5	2	2	17	5	5	3	3	9	3	10	6	5	10	2	5	12	4	4	7	23	6	6.36
35 *place*	7	8	10	5	4	8	9	5	9	17	7	5	3	5	7	5	9	5	4	12	7	9	4	3	3	7	23	9	7.46
36 *police*	5	5	4	4	2	7	10	3	9	14	5	5	10	4	3	4	5	6	7	10	11	5	12	3	6	4	10	6	6.39
37 *power*	3	10	7	6	2	3	10	5	11	8	10	5	10	4	7	4	7	12	7	12	11	3	6	3	7	9	10	9	7.18
38 *president*	5	5	7	6	2	3	10	3	11	8	5	5	10	4	3	4	8	7	7	10	11	5	12	3	7	1	10	6	6.36
39 *reason*	6	10	4	1	2	6	6	3	2	17	10	6	10	6	7	3	10	12	6	12	4	4	6	11	7	10	10	14	7.32
40 *shop*	7	3	10	5	2	8	9	5	1	8	7	5	5	5	8	5	7	5	4	12	1	9	4	3	6	7	23	9	6.54
41 *social*	3	10	4	6	6	7	10	4	9	14	6	3	10	5	9	5	7	6	5	10	11	3	6	11	5	9	23	14	7.89
42 *stand*	3	4	5	3	1	7	3	2	9	14	6	4	8	6	7	3	7	5	4	9	3	7	12	11	4	7	23	9	6.64
43 *step*	3	4	5	3	2	6	3	5	9	14	7	4	8	2	4	3	7	5	4	9	3	7	3	11	4	7	5	14	5.75
44 *street*	7	3	10	5	4	8	9	5	9	14	7	5	3	5	8	5	5	5	4	12	7	9	1	3	6	7	23	9	7.07
45 *then*	7	6	7	4	3	8	7	4	9	14	10	6	7	6	9	2	5	6	5	5	5	8	3	11	6	10	23	12	7.43
46 *understand*	6	10	4	4	6	6	6	3	11	17	10	6	6	6	3	3	10	12	6	12	4	4	12	11	7	7	10	14	7.71
47 *very*	1	4	1	4	1	1	1	3	9	14	1	1	1	1	9	4	7	1	3	9	5	5	1	11	6	7	23	12	5.21
48 *walk*	3	4	5	3	2	7	3	2	9	17	8	4	8	2	7	3	4	5	4	9	3	7	12	11	4	7	5	12	6.07
49 *war*	4	5	7	6	2	3	10	5	7	17	10	4	10	6	9	5	5	7	7	12	11	3	6	3	7	9	10	9	7.11
50 *while*	7	6	7	4	1	8	1	4	9	17	1	6	7	1	9	2	7	6	5	5	5	8	5	11	6	1	23	12	6.57

Appendix 4.3b. Number of clustered words each word made for NNS (k = 49)

	Participant																												Average
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	
1 *ago*	13	6	8	12	8	5	6	13	4	6	8	7	6	6	12	24	13	5	8	14	2	6	5	5	9	7	7	8	8.32
2 *air*	5	6	2	6	7	6	4	13	2	4	5	1	6	2	7	24	5	4	4	11	9	2	3	3	12	1	1	3	5.64
3 *all*	13	3	1	2	7	8	3	14	3	2	5	6	5	4	5	24	16	2	5	11	5	4	3	4	3	2	4	2	5.93
4 *already*	13	6	8	6	8	5	1	1	2	6	8	7	6	6	12	24	13	5	8	3	5	6	5	5	9	7	7	8	7.14
5 *also*	13	1	1	6	1	1	1	1	2	2	5	6	5	4	5	1	16	1	5	1	4	4	1	4	1	1	7	1	3.61
6 *area*	5	6	7	5	9	10	6	13	5	6	5	7	6	4	14	16	12	4	5	14	4	5	3	6	12	5	4	5	7.25
7 *arm*	10	17	2	12	14	1	6	1	2	5	9	8	5	1	1	24	16	6	6	14	9	1	3	5	13	1	3	14	7.46
8 *arrive*	13	17	7	5	9	10	7	14	2	1	10	7	6	4	12	16	12	5	1	8	6	6	3	3	9	11	1	14	7.82
9 *believe*	13	17	4	12	9	8	5	14	4	2	9	6	6	2	1	7	4	5	8	9	6	3	2	3	9	11	4	14	7.04
10 *boy*	13	5	4	6	14	1	5	14	4	5	9	2	5	5	4	7	5	5	6	9	9	14	2	4	1	6	5	5	6.21
11 *business*	13	3	4	4	14	10	5	6	2	6	9	3	5	5	12	16	16	3	4	8	7	14	7	4	13	7	1	9	7.50
12 *century*	13	1	8	12	8	6	6	14	3	2	8	7	5	7	14	24	13	4	8	9	7	2	3	5	9	1	8	8	7.68
13 *clear*	6	17	3	6	1	6	4	12	4	4	10	6	1	4	7	7	5	3	4	11	5	4	3	3	12	11	1	14	6.21
14 *close*	10	17	7	5	9	6	7	12	4	2	10	3	6	2	12	16	12	3	8	8	5	4	2	4	12	7	5	14	7.57
15 *country*	13	1	7	6	9	6	5	13	3	4	9	8	6	7	7	24	16	4	8	14	4	14	3	6	12	5	8	3	8.04
16 *cry*	13	17	4	12	9	4	4	13	3	1	10	4	5	2	4	7	4	5	8	11	5	6	2	5	9	11	8	14	7.14
17 *doctor*	6	5	2	4	14	5	5	13	4	2	9	3	5	5	4	7	16	5	6	1	9	14	7	4	13	6	5	5	6.57
18 *dream*	13	17	4	6	9	8	5	14	4	2	9	4	6	7	7	16	4	6	8	9	7	14	2	3	9	7	4	14	7.79
19 *early*	13	6	8	6	8	5	6	1	4	6	8	7	6	6	12	7	13	5	8	8	5	6	5	5	9	7	7	8	6.96
20 *figure*	10	17	1	6	14	8	4	14	2	5	9	6	5	4	12	24	13	6	2	11	6	2	2	4	9	2	3	9	7.50
21 *find*	6	17	7	6	9	8	5	14	4	2	10	6	6	4	1	16	13	5	2	9	6	3	2	2	9	11	4	14	7.18
22 *form*	10	17	1	4	7	8	5	12	1	5	5	3	6	1	14	24	12	6	2	11	6	6	2	4	13	2	3	9	7.11
23 *help*	6	17	1	12	9	1	4	13	4	2	10	4	6	2	4	24	16	6	5	11	6	6	3	5	9	11	1	14	7.57
24 *history*	13	3	8	12	8	6	6	13	3	2	8	8	5	7	14	24	13	4	8	14	7	14	3	5	9	7	8	8	8.57
25 *keep*	10	17	1	4	7	8	1	6	2	2	10	4	6	2	14	16	13	5	4	11	6	4	4	3	13	11	1	14	7.11
26 *law*	10	3	7	12	14	5	5	12	2	5	9	8	5	1	14	24	16	3	5	14	7	14	4	3	13	7	8	9	8.54
27 *lot*	1	3	7	5	9	8	3	6	3	2	5	6	5	4	5	16	16	1	5	1	5	4	3	6	3	2	4	5	5.11
28 *matter*	6	17	3	4	7	6	4	12	2	2	5	3	5	2	14	24	13	5	5	14	6	5	4	1	9	2	3	9	6.86
29 *nature*	1	1	2	6	14	6	4	13	2	4	5	3	5	2	7	24	5	4	4	11	4	2	3	3	12	5	3	3	5.64
30 *next*	13	6	8	6	8	5	6	14	4	6	8	7	6	6	12	16	12	5	8	9	4	6	2	4	9	7	7	8	7.57
31 *nothing*	13	1	1	2	7	4	4	13	3	2	5	6	5	1	5	24	13	1	5	11	1	4	1	1	3	2	4	2	5.14
32 *open*	10	17	4	5	9	6	7	12	1	2	10	3	6	2	12	16	12	3	4	8	5	4	2	4	12	11	5	14	7.36
33 *our*	13	1	1	6	14	6	5	14	5	1	1	6	5	2	1	24	16	2	6	14	1	1	7	4	13	1	1	1	6.14
34 *person*	1	5	1	4	14	10	1	12	3	5	9	2	5	5	1	7	5	2	6	9	9	14	7	4	13	6	5	5	6.07
35 *place*	5	6	7	5	9	10	1	13	5	6	3	7	6	4	7	16	12	2	5	9	4	5	3	6	12	5	4	5	6.50
36 *police*	6	5	7	4	14	5	5	12	2	5	9	8	5	5	14	24	16	5	5	14	9	14	7	3	13	6	5	5	8.29
37 *power*	13	1	7	12	14	5	6	12	3	5	9	8	5	7	14	24	16	4	8	14	7	5	3	3	13	7	8	9	8.64
38 *president*	13	5	4	12	14	5	6	12	3	5	9	3	5	7	14	24	16	5	6	14	9	14	7	4	13	6	5	5	8.75
39 *reason*	13	1	7	12	7	4	5	13	3	1	9	6	5	2	14	24	13	5	5	14	6	5	4	1	9	2	3	9	7.21
40 *shop*	5	6	4	5	9	10	7	6	5	6	9	3	5	4	12	16	12	5	4	8	9	14	2	6	12	6	4	5	7.11
41 *social*	13	1	7	6	14	6	5	12	4	1	9	8	5	2	14	24	16	2	4	14	6	14	7	4	13	7	8	1	8.11
42 *stand*	10	17	3	5	9	10	7	14	2	2	10	7	6	4	1	16	12	6	8	11	5	6	2	3	1	11	5	14	7.39
43 *step*	10	6	3	6	9	10	5	14	4	6	3	7	6	4	1	16	12	5	5	9	6	1	2	4	12	1	5	9	6.46
44 *street*	5	6	7	5	9	10	7	6	5	6	3	7	6	4	14	16	12	4	5	8	9	14	3	6	12	5	4	5	7.25
45 *then*	13	6	8	6	8	5	6	14	4	6	8	7	6	6	12	24	13	5	8	3	4	6	5	4	9	7	7	8	7.79
46 *understand*	13	17	3	4	9	6	5	12	4	2	9	6	6	4	1	24	16	5	2	1	6	3	2	2	9	11	4	14	7.14
47 *very*	13	3	1	6	1	1	3	14	4	1	5	6	5	4	5	1	4	2	5	1	4	1	3	1	1	2	4	1	3.64
48 *walk*	10	17	3	5	9	10	7	6	2	2	10	7	6	4	7	16	12	5	5	8	5	6	3	3	12	11	5	14	7.50
49 *war*	13	1	7	12	14	4	6	13	3	5	9	8	5	7	14	24	16	3	8	14	7	14	3	5	13	7	8	9	9.00
50 *while*	13	6	8	6	8	1	6	1	1	6	8	7	6	6	12	1	13	1	8	3	2	6	5	5	9	7	7	8	6.07

Appendix 4.4a. Number of connections each word made with other words for NS (*k* = 49)

	Number of connections			
	0	1 to 10	11 to 20	21 to 28
1 *ago*	8	35	6	0
2 *air*	13	34	1	1
3 *all*	7	39	3	0
4 *already*	8	35	6	0
5 *also*	11	34	4	0
6 *area*	8	38	2	1
7 *arm*	8	41	0	0
8 *arrive*	10	38	1	0
9 *believe*	11	34	3	1
10 *boy*	18	29	2	0
11 *business*	13	35	1	0
12 *century*	10	35	4	0
13 *clear*	9	38	2	0
14 *close*	10	38	1	0
15 *country*	5	42	2	0
16 *cry*	11	37	1	0
17 *doctor*	14	31	4	0
18 *dream*	12	34	3	0
19 *early*	9	35	5	0
20 *figure*	1	48	0	0
21 *find*	13	35	1	0
22 *form*	2	46	1	0
23 *help*	4	44	1	0
24 *history*	10	34	5	0
25 *keep*	1	47	1	0
26 *law*	13	32	4	0
27 *lot*	14	32	3	0
28 *matter*	2	46	1	0
29 *nature*	4	44	0	1
30 *next*	5	38	6	0
31 *nothing*	4	44	1	0
32 *open*	15	33	1	0
33 *our*	2	47	0	0
34 *person*	7	40	2	0
35 *place*	9	36	2	2
36 *police*	15	31	3	0
37 *power*	8	37	4	0
38 *president*	17	29	3	0
39 *reason*	8	37	4	0
40 *shop*	11	33	5	0
41 *social*	2	46	1	0
42 *stand*	0	47	2	0
43 *step*	10	37	2	0
44 *street*	7	39	2	1
45 *then*	5	38	6	0
46 *understand*	7	39	2	1
47 *very*	7	41	1	0
48 *walk*	4	42	3	0
49 *war*	8	36	5	0
50 *while*	12	32	5	0

Appendix 4.4b. Number of connections each word made with other words for NNS (k = 49)

	Number of connections			
	0	1 to 10	11 to 20	21 to 28
1 *ago*	7	35	4	3
2 *air*	10	37	2	0
3 *all*	8	38	3	0
4 *already*	11	33	1	4
5 *also*	22	27	0	0
6 *area*	11	35	1	2
7 *arm*	5	44	0	0
8 *arrive*	9	37	3	0
9 *believe*	10	36	3	0
10 *boy*	10	36	3	0
11 *business*	5	41	3	0
12 *century*	7	39	3	0
13 *clear*	10	37	2	0
14 *close*	9	36	3	1
15 *country*	6	41	2	0
16 *cry*	5	43	1	0
17 *doctor*	12	33	4	0
18 *dream*	5	42	2	0
19 *early*	16	28	1	4
20 *figure*	4	44	1	0
21 *find*	9	36	4	0
22 *form*	4	44	1	0
23 *help*	7	41	1	0
24 *history*	8	37	4	0
25 *keep*	4	44	1	0
26 *law*	7	36	6	0
27 *lot*	12	35	2	0
28 *matter*	6	42	1	0
29 *nature*	9	38	2	0
30 *next*	14	30	4	1
31 *nothing*	11	37	1	0
32 *open*	11	35	2	1
33 *our*	6	42	1	0
34 *person*	10	35	4	0
35 *place*	14	31	3	1
36 *police*	10	32	7	0
37 *power*	7	36	6	0
38 *president*	10	31	8	0
39 *reason*	5	43	1	0
40 *shop*	11	34	4	0
41 *social*	8	33	8	0
42 *stand*	10	34	5	0
43 *step*	10	36	3	0
44 *street*	13	33	2	1
45 *then*	8	35	1	5
46 *understand*	4	43	2	0
47 *very*	26	21	2	0
48 *walk*	18	26	5	0
49 *war*	10	32	7	0
50 *while*	26	18	2	3

Appendix 4.5a. Linked words of high frequency (count of 15 or more) in clusters made by NS (k = 50) (n = 28)

Cluster category	*Word*	No.	*Linked word* (frequency)
TIME	*ago*	5	*already* (19), *century* (16), *early* (17), *next* (15), *then* (18)
	already	4	*ago* (19), *next* (16), *then* (18), *while* (18)
	next	4	*ago* (15), *already* (16), *then* (16), *while* (16)
	then	4	*ago* (18), *already* (18), *next* (16), *while* (17)
	while	3	*already* (18), *next* (16), *then* (17)
	early	1	*ago* (17)
POWER	*police*	3	*doctor* (18), *law* (15), *president* (17)
	doctor	1	*police* (18)
	law	1	*police* (15)
	power	1	*war* (17)
	president	1	*police* (17)
THOUGHT	*believe*	3	*dream* (18), *reason* (16), *understand* (22)
	understand	3	*believe* (22), *dream* (16), *reason* (20)
	reason	2	*believe*(16), *understand* (20)
	dream	2	*believe* (18), *understand* (16)
PLACE	*shop*	4	*area* (17), *business* (15), *place* (19), *street* (19)
	area	3	*place* (22), *shop* (17), *street* (19)
	street	3	*area* (19), *place* (21), *shop* (19)
	place	2	*area* (22), *street* (21)
MOVEMENT	*walk*	2	*stand* (20), *step* (20)
	stand	1	*walk* (20)
	step	1	*walk* (20)
HISTORY	*century*	2	*ago* (16), *history* (17)
	history	2	*century* (17), *war* (17)
	war	2	*history* (17), *power* (17)
PERSON	*boy*	1	*person* (19)
	person	1	*boy* (19)
OPEN/CLOSE	*close*	1	*open* (19)
	open	1	*close* (19)
NATURE	*air*	1	*nature* (21)
	nature	1	*air* (21)
BUSINESS	*business*	1	*shop* (15)
ALSO/WHILE	*also*	1	*while* (15)

Note. No. = number of links the word produced. Number in parentheses shows the number of participants.

Appendix 4.5b. Linked words of high frequency (count of 15 or more) in clusters made by NNS (k = 50) (n = 28)

Cluster category	*Word*	No.	*Linked word* (frequency)
TIME	*already*	5	*ago* (21), *early* (23), *next* (18), *then* (23), *while* (21)
	early	5	*ago* (21), *already* (23), *next* (19), *then* (22), *while* (21)
	next	5	*ago* (19), *already* (18), *early* (19), *then* (23), *while* (17)
	then	5	*ago* (22), *already* (23), *early* (22), *next* (23), *while* (21)
	while	5	*ago* (20), *already* (21), *early* (21), *next* (17), *then* (21)
	ago	4	*already* (21), *next* (19), *then* (22), *while* (20)
POWER	*law*	3	*police* (17), *power* (18), *war* (17)
	police	3	*doctor* (16), *law* (17), *president* (18)
	war	3	*history* (17), *law* (17), *power* (19)
	power	2	*law* (18), *war* (19)
	president	2	*doctor* (15), *police* (18)
	social	1	*law* (15)
PLACE	*street*	3	*area* (21), *place* (20), *shop* (19)
	area	2	*place* (23), *street* (21)
	place	2	*area* (23), *street* (20)
	shop	1	*street* (19)
THOUGHT	*believe*	2	*dream* (17), *find* (15)
	find	2	*believe* (15), *understand* (17)
	dream	1	*believe* (17)
	under-stand	1	*find* (17)
PERSON	*doctor*	4	*boy* (15), *person* (15), *police* (16), *president* (15)
	boy	2	*doctor* (15), *person* (17)
	person	2	*boy* (17), *doctor* (15)
NOTHING/ALL	*all*	2	*lot* (16), *nothing* (16)
	nothing	1	*all* (16)
	lot	1	*all* (16)
MOVEMENT	*arrive*	2	*stand* (16), *walk* (17)
	stand	2	*arrive* (16), *walk* (19)
	walk	2	*arrive* (17), *stand* (19)
REASON/MATTER	*matter*	1	*reason* (15)
	reason	1	*matter* (15)
OPEN/CLOSE	*close*	1	*open* (22)

	open	1	*close* (22)
NATURE	*air*	1	*nature* (20)
	nature	1	*air* (20)
HELP	*cry*	1	*help* (15)
	help	1	*cry* (15)
HISTORY	*history*	2	*century* (18), *war* (17)
	century	1	*history* (18)

Note. No. = number of links the word produced. Number in parentheses shows the number of participants.

Appendix 5.1a: Number of times words were associated with other words by NS (n = 30) (k = 50)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	*accept*	30	4	15	8	8	6	1	8	1	2	1	13	8	1	10	2	3	4	3	2	2	4	6	5	8	3	3	4	2	2	2	5	3	2	2	2	3	3	11	14	5	1	1	1	7	2	2		2	2
2	*add*	4	30	4	4	4	3	4	3	10	3	2	3	3	8	3	2	4	3	5		2	4	1	4	3	9		4	7	3	7	1		4	5	2		2	2	2	1	2	1	2	3	1	2	1	1	2
3	*agree*	15	4	30	13	9	6		8		2	1	13	10		13	4		9	5	3	4	2	4	1	5	2	6	2		3	5	1	4	2	5	1	2	2	4	14	3	2	4	2	8	1	4		4	2
4	*argue*	8	4	13	30	14	10	1	8		1	2	9	13	1	9	8	2	18	1		8	4		2	3	2	4	2	2	3	2	3	5	4	1	2	2	4	3	13	4	1	8	1	11	3	8	1	2	5
5	*ask*	8	4	9	14	30	2	1	5	2	2	1	6	8	3	6	9	6	17	1		9	3	2	2	3	4	8	4	4	3	1	10	11	1	1	1	1	8	4	4	6	1	12	1	10	5	12		1	7
6	*avoid*	6	3	6	10	2	30	3	4		1	3	6	4		6			2	6	2		3	4	2	3	3	1			2	5			8	4	2	8		2	15		1		2	1		1	1	8	
7	*begin*	1	4		1	1	3	30	2	5	1	3	1	1	6	1	3	9	2	12	3	3	4	4	1	1	8	1	2	9	9	7	3	3	1	7	6	2	2			3		2	2	2	2	1	2	4	2
8	*believe*	8	3	8	8	5	4	2	30	1		2	10	15	4	9	3	6	6		6	3	3	3	1	1	3	3	18	4	2	3	4	3	1	2	1	3	4	1	5	13		3		6	2	5		1	4
9	*build*	1	10			2		5	1	30	1	1			18			9			1		1	2	2		17		6	15		1	4	2	2		5	2				2	1		1		1		1		
10	*buy*	2	3	2	1	2	1	1		1	30	3	2	2		2		1		3	1		3		14	4	3	3		1		1	1	1	2	1		2	2	8	2		21	1	25	1		1	1	1	1
11	*carry*	1	2	1	2	1	3	3	2	1	3	30	1	1		1				6			4	2	4	2	2	1	1	1		2		1	4	3		5	1	3	1		3	1	5			1	13	1	
12	*choose*	13	3	13	9	6	6	1	10		2	1	30	15	1	27	2	1	5	4	2	2	4	3	3	2	1	3	4	1	2	2	2	1	2	3	3	2	1	2	10	4	2	1	2	7	1	2		3	1
13	*consider*	8	3	10	13	8	4	1	15		2	1	15	30	1	16	5	4	10	2	2	5	6		3	2	3	4	10	4	2	1	6	4	1	2	2	1	5	2	7	12	2	4	2	9	3	4		2	5
14	*create*	1	8		1	3		6	4	18			1	1	30	1	1	13	1		2	1	5			1	10		13	13	2	1	7	1			9		3			3				1	3	1		1	3
15	*decide*	10	3	13	9	6	6	1	9		2	1	27	16	1	30	2	2	5	4	1	2	4	4	3	1	2	4	3	2	2	2	2	2	2	3	2	2	2	1	10	5	2	1	2	6	1	2		4	2
16	*describe*	2	2	4	8	9		3	3				2	5	1	2	30	4	16		2	26	2	1		4		6	2	3	14	2	6	7	1	2	5		7	1	1	6	2	15		16	19	21	1	1	10
17	*discover*	3	4		2	6		9	6	9	1		1	4	13	2	4	30	5	1	3	4	14	2	1		11	4	8	15	4	2	13	7	1	2	4		7	1		8		3		5	5	4		1	6
18	*discuss*	4	3	9	18	17	2	2	6				5	10	1	5	16	5	30		2	15	2	1		1		7	2	4	6	2	6	8		2	3	1	8	1	3	4		15		18	9	14		1	9
19	*enter*	3	5	5	1	1	6	12			3	6	4	2		4		1		30			3	7	2	2	3	3		1	6	14		1	3	15	2	3	1	1	3		3	1	5				4	7	
20	*expect*	2		3			2	3	6	1	1		2	2	2	1	2	3	2		30	2	2	3	2	3	1	1	4	3	2	3	1	1	1	3	8		1	3	1	2	1	2	1	4	2	1		14	1
21	*explain*	2	2	4	8	9		3	3				2	5	1	2	26	4	15		2	30	2	1		4		6	1	3	15	2	6	7	1	2	5		7	1	1	5	2	15		16	22	23	1	1	10
22	*find*	4	4	2	4	3	3	4	3	1	3	4	4	6	5	4	2	14	2	3	2	2	30	3	6	2	6		3	5	1	2	6	2	11	1	3	1	4	2	2	5	1	1	3	3	2	1	1	2	4
23	*follow*	6	1	4		2	4	4	3	2		2	3			4	1	2	1	7	3	1	3	30			4	4		4	5	12	3	5	2	11		5	2	1	2	2		2	1	2	1	1	1	7	2
24	*get*	5	4	1	2	2	2	1	1	2	14	4	3	3		3		1		2	2		6		30	10	3	2		1		1	1	1	5			2	2	18	2		8	1	11	1		1	3	2	1
25	*give*	8	3	5	3	3	3	1	1		4	2	2	2	1	1	4		1	2	3	4	2		10	30	3	2	1	2	4	2	1	1	4	2	3	1	2	17	5	1	8	4	5	3	4	5	3	1	3
26	*grow*	3	9	2	2	4	3	8	3	17	3	2	1	3	10	2		11		3	1		6	4	3	3	30		6	19	1	4	8	3	2	2	1	3	2	3	2	5	2		3		1			2	2
27	*hear*	3		6	4	8	1	1	3		3	1	3	4		4	6	4	7	3	1	6		4	2	2		30	1	1	3	3	7	26	1	4		1	13	2	4	3	2	17	2	6	3	9	1	2	12
28	*imagine*	4	4	2	2	4		2	18	6		1	4	10	13	3	2	8	2		4	1	3			1	6	1	30	7	2		8	2		1	5	1	6	1	1	13		1		4	2	2		1	5
29	*improve*	2	7		2	4		9	4	15	1	1	1	4	13	2	3	15	4	1	3	3	5	4	1	2	19	1	7	30	3	1	13	5		1	7	2	5			6		2	1	4	5	2	1	1	5
30	*introduce*	2	3	3	3	3	2	9	2				2	2	2	2	14	4	6	6	2	15	1	5		4	1	3	2	3	30	10	3	2	2	11	6	1	4	1	1	4	2	8		12	12	11	1	3	4
31	*join*	2	7	5	2	1	5	7	3	1	1	2	2	1	1	2	2	2	2	14	3	2	2	12	1	2	4	3		1	10	30	1	2	2	24	1	3	2	3	2	1		2	1	2	1	3		6	2
32	*learn*	5	1	1	3	10		3	4	4	1		2	6	7	2	6	13	6		1	6	6	3	1	1	8	7	8	13	3	1	30	11		1	5		16	1	1	13		3		6	13	5			13
33	*listen*	3		4	5	11		3	3	2	1	1	1	4	1	2	7	7	8	1	1	7	2	5	1	1	3	26	2	5	2	2	11	30		2	1		14	2	2	5		17		7	4	8	1	1	13
34	*lose*	2	4	2	4	1	8	1	1	2	2	4	2	1		2	1	1		3	1	1	11	2	5	4	2	1			2	2			30	1	1	4		1	5		5	1	3		1	2	3	4	1
35	*meet*	2	5	5	1	1	4	7	2		1	3	3	2		3	2	2	2	15	3	2	1	11		2	2	4	1	1	11	24	1	2	1	30	1	3	3	2	3	1	2	2	3	2	1	2	1	6	2
36	*prepare*	2	2	1	2	1	2	6	1	5			3	2	9	2	5	4	3	2	8	5	3			3	1		5	7	6	1	5	1	1	1	30	2	3			2		3		7	7	2		5	2
37	*protect*	3		2	2	1	8	2	3	2	2	5	2	1		2			1	3			1	5	2	1	3	1	1	2	1	3			4	3	2	30		1	3		1		3			1	2	1	
38	*read*	3	2	2	4	8		2	4		2	1	1	5	3	2	7	7	8	1	1	7	4	2	2	2	2	13	6	5	4	2	16	14		3	3		30	2	1	6	2	10	2	8	9	8	1		25
39	*receive*	11	2	4	3	4	2		1		8	3	2	2		1	1	1	1	1	3	1	2	1	18	17	3	2	1		1	3	1	2	1	2		1	2	30	6	1	4	1	5	2		1	3	2	1
40	*refuse*	14	2	14	13	4	15		5		2	1	10	7		10	1		3	3	1	1	2	2	2	5	2	4	1		1	2	1	2	5	3		3	1	6	30	2	2	1	2	2	1	1	1	6	1
41	*remember*	5	1	3	4	6		3	13	2			4	12	3	5	6	8	4		2	5	5	2		1	5	3	13	6	4	1	13	5		1	2		6	1	2	30		2		4	5	4			5
42	*sell*	1	2	2	1	1	1			1	21	3	2	2		2	2			3	1	2	1		8	8	2	2			2				5	2		1	2	4	2		30	3	21	1	2	3	2	1	2
43	*speak*	1	1	4	8	12		2	3		1	1	1	4		1	15	3	15	1	2	15	1	2	1	4		17	1	2	8	2	3	17	1	2	3		10	1	1	2	3	30	1	12	9	17	2	1	13
44	*spend*	1	2	2	1	1	2	2		1	25	5	2	2		2				5	1		3	1	11	5	3	2		1		1			3	3		3	2	5	2		21	1	30			1	2	1	1
45	*suggest*	7	3	8	11	10	1	2	6		1		7	9	1	6	16	5	18		4	16	3	2	1	3		6	4	4	12	2	6	7		2	7		8	2	2	4	1	12		30	12	12		1	6
46	*teach*	2	1	1	3	5		2	2	1			1	3	3	1	19	5	9		2	22	2	1		4	1	3	2	5	12	1	13	4	1	1	7		9		1	5	2	9		12	30	16	1	1	10
47	*tell*	2	2	4	8	12	1	1	5		1	1	2	4	1	2	21	4	14		1	23	1	1	1	5		9	2	2	11	3	5	8	2	2	2	1	8	1	1	4	3	17	1	12	16	30	1		11
48	*throw*		1		1		1	2		1	1	13					1			4		1	1	1	3	3		1		1	1			1	3	1		2	1	3	1		2	2	2		1	1	30		1
49	*wait*	2	1	4	2	1	8	4	1		1	1	3	2	1	4	1	1	1	7	14	1	2	7	2	1	2	2	1	1	3	6		1	4	6	5	1		2	6		1	1	1	1	1			30	
50	*write*	2	2	2	5	7		2	4		1		1	5	3	2	10	6	9		1	10	4	2	1	3	2	12	5	5	4	2	13	13	1	2	2		25	1	1	5	2	13	1	6	10	11	1		30

Note. Blank cells equal "0".

Appendix 5.1b: Number of times words were associated with other words by NNS (n = 30) (k = 50)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	*accept*	30	2	16	6	2	11	1	7	1	3	1	7	6	2	8	1	2	6	2	5	2	2	6	3	2
2	*add*	2	30	3	3	2	4	4		7	1	5	2	1	3	1	3		2	3	1	3		4	2	1
3	*agree*	16	3	30	15	5	6	1	9		1		5	9	4	12	7	3	13		6	6	1	5		1
4	*argue*	6	3	15	30	10	1	1	4		1		2	5	3	7	9	2	26		1	11		2		1
5	*ask*	2	2	5	10	30		2	1	1	1		1	2	2	2	9	3	8	2	1	14	4	2		1
6	*avoid*	11	4	6	1		30	3	1	3		2	2	2	1	3			2	5	2			8	1	
7	*begin*	1	4	1	1	2	3	30		6		2	3		4	1	3	3		13	1	2	4	7	1	
8	*believe*	7		9	4	1	1		30		1		5	20	4	13	4	6	4		18	2	4	1		1
9	*build*	1	7			1	3	6		30	1	6	1		15		3	8	1	5	2	1	6	5	3	1
10	*buy*	3	1	1	1	1			1	1	30	4	8		1			2	1		1		3		11	11
11	*carry*	1	5				2	2		6	4	30	6		2			1	1	3	2			4	7	6
12	*choose*	7	2	5	2	1	2	3	5	1	8	6	30	6	7	12	4	6	1	1	5	2	6	3	6	3
13	*consider*	6	1	9	5	2	2		20				6	30	5	19	4	7	4		14	6	5	2		
14	*create*	2	3	4	3	2	1	4	4	15	1	2	7	5	30	9	5	16	2	1	6	2	8	1	2	1
15	*decide*	8	1	12	7	2	3	1	13				12	19	9	30	6	8	5		9	3	6	1		
16	*describe*	1	3	7	9	9		3	4	3			4	4	5	6	30	6	10	2	3	18	3	3		1
17	*discover*	2		3	2	3		3	6	8	2	1	6	7	16	8	6	30	2	3	8	3	18	3	3	1
18	*discuss*	6	2	13	26	8	2		4	1	1	1	1	4	2	5	10	2	30	1	1	11		3	1	1
19	*enter*	2	3			2	5	13		5		3	1		1		2	3	1	30	1	2	3	8	2	
20	*expect*	5	1	6	1	1	2	1	18	2	1	2	5	14	6	9	3	8	1	1	30	1	5	1	2	1
21	*explain*	2	3	6	11	14		2	2	1			2	6	2	3	18	3	11	2	1	30	3	3		1
22	*find*	2		1		4		4	4	6	3		6	5	8	6	3	18		3	5	3	30	3	4	3
23	*follow*	6	4	5	2	2	8	7	1	5		4	3	2	1	1	3	3	3	8	1	3	3	30	1	
24	*get*	3	2				1	1		3	11	7	6		2			3	1	2	2		4	1	30	20
25	*give*	2	1	1	1	1			1	1	11	6	3		1		1	1	1		1	1	3		20	30
26	*grow*		4			2		5	1	8	1	4	2	1	10			6		2	3		4	1	3	2
27	*hear*	5	1	5	4	7	1	1	3				1	4	1	4	3	1	4	1	2	5	4			1
28	*imagine*	5		6	2		1	1	23	3			4	18	8	13	4	8	2		19	1	6		1	
29	*improve*	2	3	3	2	7		4	3	5	2	1	5	5	14	5	4	10	1	1	7	3	6		3	1
30	*introduce*	3		3	8	8	1	4	2	2	1	2	1		1	1	10	2	9	2	3	12	2	3	1	3
31	*join*	4	6	2	1	1	5	8	2	1	2	3	4		1			1	2	12	2			8	3	3
32	*learn*	2	3	1	5	12	1	5	1	5	1	1	3	3	6	3	7	8	4	3	4	7	9	3	2	
33	*listen*	2	1	3	6	11			1				1	2	1	1	5	2	6		1	8	4			1
34	*lose*	4	2				6	3		3	5	2	2		3	1		2		1	2		6	2	11	13
35	*meet*	5	1	3	2	3	6	4	5	2	1	2	2	1		2	2	1	3	6	3	1	4	7	1	1
36	*prepare*	2	4	1	1	2	1	5	1	5	2	6	3	2	5		1	4	1	2	3	3	2	4	2	2
37	*protect*	4	4	3	1	1	5	3	1	4	1	4	3		4	1	1	2		3	1	1	1	5	3	2
38	*read*	1	2	1	2	9	1	4	1	3			2	2	2	1	6	3	2	3	2	5	5	3		1
39	*receive*	9	2	3	2	1	5	2	3	2	13	7	8	1	2	2	1	3	3	4	2	1	2	4	16	15
40	*refuse*	19	2	13	7	3	16		2				3	2	1	5	3		6	1	2	3		5		
41	*remember*	4	1	4		1	1		18				4	17	4	13	3	5			12	3	6			
42	*sell*	1		1	1	1			1	1	27	5	6		1			1	1		1		2		9	14
43	*speak*	1	1	2	6	11								1			6	1	6		1	12	2			2
44	*spend*					1		3	1	2	16	2	5	1	2		1	3			2		4	2	7	9
45	*suggest*	6	1	11	15	10	3		6		1	1	2	8	2	6	10	4	15		4	14	1	3		2
46	*teach*	1	2	2	6	16	1	2	1	1	1	1	1	1	2	1	9	2	5	1	1	14	3	2		3
47	*tell*	3	1	6	12	13			4		1	1		3	1	3	12	2	13		2	15	1			3
48	*throw*	1	5				6	4		5	1	16	3						1	5	1			5	5	4
49	*wait*	1	1	1		3	3	6	3	2	1	2	4	1	1		2	3		5	3	1	3	8	1	1
50	*write*	1	5		2	8	2	5		6		2	1	1	2		7	3	2	4	2	4	4	4		1

		26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	*accept*		5	5	2	3	4	2	2	4	5	2	4	1	9	19	4	1	1		6	1	3	1	1	1
2	*add*	4	1		3		6	3	1	2	1	4	4	2	2	2	1		1		1	2	1	5	1	5
3	*agree*		5	6	3	3	2	1	3		3	1	3	1	3	13	4	1	2		11	2	6		1	
4	*argue*		4	2	2	8	1	5	6		2	1	1	2	2	7		1	6		15	6	12			2
5	*ask*	2	7		7	8	1	12	11		3	2	1	9	1	3	1	1	11	1	10	16	13		3	8
6	*avoid*		1	1		1	5	1		6	6	1	5	1	5	16	1				3	1		6	3	2
7	*begin*	5	1	1	4	4	8	5		3	4	5	3	4	2					3		2		4	6	5
8	*believe*	1	3	23	3	2	2	1	1		5	1	1	1	3	2	18	1		1	6	1	4		3	
9	*build*	8		3	5	2	1	5		3	2	5	4	3	2			1		2		1		5	2	6
10	*buy*	1			2	1	2	1		5	1	2	1		13			27		16	1	1	1	1	1	
11	*carry*	4			1	2	3	1		2	2	6	4		7			5		2	1	1	1	16	2	2
12	*choose*	2	1	4	5	1	4	3	1	2	2	3	3	2	8	3	4	6		5	2	1		3	4	1
13	*consider*	1	4	18	5			3	2		1	2		2	1	2	17		1	1	8	1	3		1	1
14	*create*	10	1	8	14	1	1	6	1	3		5	4	2	2	1	4	1		2	2	2	1		1	2
15	*decide*		4	13	5	1		3	1	1	2		1	1	2	5	13				6	1	3			
16	*describe*		3	4	4	10		7	5		2	1	1	6	1	3	3		6	1	10	9	12		2	7
17	*discover*	6	1	8	10	2	1	8	2	2	1	4	2	3	3		5	1	1	3	4	2	2		3	3
18	*discuss*		4	2	1	9	2	4	6		3	1		2	3	6		1	6		15	5	13	1		2
19	*enter*	2	1		1	2	12	3		1	6	2	3	3	4	1						1		5	5	4
20	*expect*	3	2	19	7	3	2	4	1	2	3	3	1	2	2	2	12	1	1	2	4	1	2	1	3	2
21	*explain*		5	1	3	12		7	8		1	3	1	5	1	3	3		12		14	14	15		1	4
22	*find*	4	4	6	6	2		9	4	6	4	2	1	5	2		6	2	2	4	1	3	1		3	4
23	*follow*	1				3	8	3		2	7	4	5	3	4	5				2	3	2		5	8	4
24	*get*	3		1	3	1	3	2		11	1	2	3		16			9		7				5	1	
25	*give*	2	1		1	3	3		1	13	1	2	2	1	15			14	2	9	2	3	3	4	1	1
26	*grow*	30		3	12		5	4		3	3	6	3	1	2			1		4	1	2	1	1	5	1
27	*hear*		30	3	2	2		6	23		2	2		17		4	8		18		3	4	10	1	1	16
28	*imagine*	3	3	30	6	1		3	1		2	1		1		2	19			1	4		2		1	
29	*improve*	12	2	6	30	2	2	11	5	1		4	2	5	2		5	1	4	1	2	7	2			4
30	*introduce*		2	1	2	30	5	5	5		11	1	1	4	2	3	1	2	9	1	11	11	12	2	7	3
31	*join*	5			2	5	30			1	11	2	2		6	3		2		1	1	2	1	3	7	
32	*learn*	4	6	3	11	5		30	8	1		5	2	12	2	1	5		7	1	3	17	5	1	1	13
33	*listen*		23	1	5	5		8	30		1	2		18		4	5		23		7	7	13	1		17
34	*lose*	3			1		1	1		30	1	3	4		10	6		7	1	9	1			6	2	1
35	*meet*	3	2	2		11	11		1	1	30		3	1	4	3	1	1		1	2	2	2	3	18	1
36	*prepare*	6	2	1	4	1	2	5	2	3		30	3	4	3			2	2	5	3	4	2	2	4	5
37	*protect*	3			2	1	2	2		4	3	3	30	1	4	3		1				1		6	3	2
38	*read*	1	17	1	5	4		12	18		1	4	1	30			4		18	1	2	9	8	2	3	26
39	*receive*	2			2	2	6	2		10	4	3	4		30	3		10		5	3	1	2	5	3	1
40	*refuse*		4	2		3	3	1	4	6	3		3		3	30	1		3		7	2	4	2		
41	*remember*		8	19	5	1		5	5		1			4		1	30		3			2				3
42	*sell*	1			1	2	2			7	1	2	1		10			30	1	17	2	2	2	3	1	
43	*speak*		18		4	9		7	23	1		2		18		3	3	1	30	1	8	10	18	2		16
44	*spend*	4		1	1	1	1	1		9	1	5		1	5			17	1	30	1	1	1	1	7	1
45	*suggest*	1	3	4	2	11	1	3	7	1	2	3		2	3	7		2	8	1	30	10	13	1	1	2
46	*teach*	2	4		7	11	2	17	7		2	4	1	9	1	2	2	2	10	1	10	30	10	1	1	8
47	*tell*	1	10	2	2	12	1	5	13		2	2		8	2	4		2	18	1	13	10	30	1	1	7
48	*throw*	1	1			2	3	1	1	6	3	2	6	2	5	2		3	2	1	1	1	1	30	2	4
49	*wait*	5	1	1		7	7	1		2	18	4	3	3	3			1		7	1	1	1	2	30	3
50	*write*	1	16		4	3		13	17	1	1	5	2	26	1		3		16	1	2	8	7	4	3	30

Note. Blank cells equal "0".

Appendix 5.2a: Presence of pairs of "native-like" links that were produced by half or more of NS participants (*n* = 30)

Participant / Word pair	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
accept-agree	0	1	1	1	0	1	0	0	0	1	0	1	1	0	0	0	0	1	0	1	0	1	0	1	1	1	0	0	1	1
argue-discuss	0	1	0	0	1	0	1	1	0	0	1	1	1	0	0	1	1	0	1	0	1	1	1	0	1	1	1	1	0	1
ask-discuss	0	0	0	1	1	1	1	0	1	1	1	0	1	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1	0	0
avoid-refuse	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	1	0	1	1	1	0	0	0	1	1	1	0	1	0	1
believe-consider	0	1	1	0	1	0	1	1	0	0	1	0	0	1	0	0	1	0	1	1	1	0	0	1	1	1	0	1	0	0
believe-imagine	1	0	0	1	0	0	1	1	0	1	1	1	0	1	0	0	1	0	1	1	1	0	1	1	0	0	1	1	1	1
build-create	0	1	0	0	1	1	0	1	0	1	1	0	0	1	0	0	0	1	0	1	1	1	1	1	1	1	0	1	1	1
build-grow	0	1	0	1	1	0	1	0	1	1	0	1	0	1	1	1	1	0	1	0	0	1	0	1	1	1	0	0	0	1
build-improve	0	1	0	0	1	0	0	1	1	1	0	1	0	0	1	1	1	0	1	0	1	1	0	1	1	1	0	0	0	0
buy-sell	1	1	1	1	1	0	1	0	1	1	0	0	1	0	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1
buy-spend	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	1	1	1
choose-consider	1	1	1	1	1	0	1	0	0	0	1	1	0	0	1	0	0	1	0	0	1	0	0	1	1	1	0	0	1	0
choose-decide	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1
consider-decide	1	1	1	1	1	0	1	0	0	1	1	1	0	0	1	0	0	1	0	0	0	0	1	1	1	1	0	0	1	0
describe-discuss	1	0	1	0	1	0	1	0	1	1	0	0	0	0	1	0	1	1	1	1	0	1	0	1	0	1	1	0	0	1
describe-explain	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	0	1	1
describe-speak	0	0	1	0	1	0	0	1	1	0	1	0	0	0	1	0	0	1	1	1	0	1	1	0	0	1	1	0	1	1
describe-suggest	1	1	0	0	1	1	1	0	0	0	1	0	0	0	1	0	1	1	1	1	0	1	1	0	0	0	1	1	0	1
describe-teach	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	0	0	0	0	1	1
describe-tell	0	0	0	0	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0
discover-improve	0	1	0	0	1	1	1	0	0	1	0	1	0	0	1	0	0	1	0	0	1	1	1	0	1	1	1	0	0	1
discuss-explain	1	0	0	0	1	0	1	0	1	1	0	0	0	0	1	0	1	1	1	1	0	1	0	1	0	1	1	0	0	1
discuss-speak	0	0	1	1	1	1	0	0	1	0	0	0	0	0	1	0	0	1	1	1	1	1	0	0	0	1	1	1	0	1
discuss-suggest	1	0	0	1	1	0	1	0	0	0	0	1	1	1	1	0	1	1	1	1	1	1	0	0	1	0	1	0	1	1
enter-meet	0	0	1	1	1	0	0	1	0	1	0	1	0	0	1	1	1	0	1	0	1	1	0	0	0	1	0	0	1	1
explain-introduce	0	1	1	1	0	1	1	1	0	0	1	0	0	1	1	0	0	1	1	1	1	0	1	0	0	0	1	0	0	0
explain-speak	0	0	0	0	1	0	0	1	1	0	1	0	0	0	1	0	0	1	1	1	0	1	1	0	1	1	1	0	1	1
explain-suggest	1	1	1	0	1	1	1	0	0	0	1	0	0	0	1	0	1	1	1	1	0	1	1	0	0	0	1	0	0	1
explain-teach	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	1	1	0	0	0	1	1	1
explain-tell	0	0	1	0	1	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0
get-receive	0	0	1	0	0	1	0	1	1	1	0	1	0	1	1	0	1	1	1	1	1	1	0	1	1	0	0	1	0	1
give-receive	1	0	0	0	1	0	0	0	1	0	1	0	1	0	1	1	1	1	1	0	0	1	0	1	1	1	1	0	1	1
grow-improve	1	1	0	0	1	1	0	0	1	1	0	1	0	0	1	1	1	0	1	1	0	1	1	1	1	1	1	1	0	0
hear-listen	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0
hear-speak	1	1	1	1	1	0	1	0	0	1	0	1	1	0	0	1	1	1	0	0	0	1	0	1	1	1	1	0	0	0
join-meet	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	0	0	1	0	1	1	1	1	1	1	1
learn-read	1	0	0	0	1	0	1	1	1	1	0	0	1	0	0	0	1	1	1	1	0	1	1	0	0	1	0	0	1	1
listen-speak	1	0	1	1	1	0	1	0	0	1	0	1	1	0	0	1	1	1	0	0	0	1	0	1	1	1	1	0	0	1
read-write	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0	1
sell-spend	1	1	1	1	1	1	1	0	1	1	0	0	1	0	1	1	1	1	1	0	1	0	1	1	1	1	0	0	0	1
speak-tell	1	1	0	1	1	1	0	1	1	0	1	0	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	0

Note. 1 = presence; 0 = absence.

Appendix 5.2b: Presence of pairs of "native-like" links that were produced by half or more of NNS participants (*n* = 30)

Participant / Word pair	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
accept-agree	1	0	0	1	1	0	0	0	1	1	1	1	1	0	0	0	1	1	0	1	1	0	0	0	0	1	1	0	1	1
argue-discuss	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1
ask-discuss	0	0	0	0	0	0	0	1	0	1	1	1	0	0	0	0	1	0	0	1	1	0	0	0	1	0	0	0	0	0
avoid-refuse	0	0	1	1	1	1	0	0	1	0	1	1	0	0	1	1	0	1	0	0	1	0	1	1	1	0	1	1	0	0
believe-consider	0	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	0	0	1	1	1	1	0	0	0	1	1
believe-imagine	1	0	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	1	0	1	1	0	1	1	1	1
build-create	1	1	0	0	1	1	1	0	1	1	0	0	0	0	0	0	0	1	0	1	0	1	0	0	1	1	1	0	1	1
build-grow	1	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	1	0	0	0
build-improve	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1
buy-sell	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1
buy-spend	1	0	0	1	0	0	0	1	0	0	1	0	1	0	1	1	1	0	1	0	1	1	1	0	0	1	1	1	0	1
choose-consider	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0
choose-decide	0	0	1	0	0	0	1	1	1	1	1	0	0	1	0	1	0	0	1	1	0	0	0	0	0	0	1	0	0	1
consider-decide	0	1	0	1	1	0	0	1	0	1	1	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0
describe-discuss	1	0	0	0	0	0	1	1	0	1	0	0	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	1	0	0
describe-explain	0	0	0	0	0	1	0	1	0	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1
describe-speak	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0
describe-suggest	1	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	1	0	0	0	0	1	1	0	1	0	1	0	0	1
describe-teach	0	0	0	0	0	0	0	1	1	1	0	0	0	1	0	1	1	0	1	0	1	0	0	0	0	0	1	0	0	0
describe-tell	1	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	1	1	1	0	0	1	1	1	1	0	0	1	0	0
discover-improve	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	1	1	0	0	1	1	0	1	0	0	0
discuss-explain	0	0	1	0	0	0	0	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	0	1	0	0
discuss-speak	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0
discuss-suggest	1	0	1	1	0	0	0	1	1	1	1	0	1	0	1	0	0	0	1	1	1	0	1	0	1	1	0	0	0	0
enter-meet	0	0	0	1	0	1	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
explain-introduce	0	0	0	0	0	0	0	1	0	1	1	0	0	0	1	0	1	1	1	0	1	0	0	0	1	1	1	0	1	0
explain-speak	1	0	0	1	1	0	0	1	0	1	1	0	0	0	0	0	1	1	1	0	0	1	0	0	0	1	0	0	1	0
explain-suggest	0	0	1	0	1	0	0	1	1	1	1	1	0	0	0	0	1	0	0	0	0	1	1	0	1	1	1	0	0	1
explain-teach	1	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1	1	0	1	0	1	0	0	0	0	0	1	0	0	0
explain-tell	0	0	0	1	1	0	0	1	0	1	1	0	0	0	0	0	1	1	1	0	0	1	1	1	1	1	0	1	1	0
get-receive	1	0	0	0	1	1	1	1	0	0	1	0	1	1	1	0	1	0	0	0	0	1	0	1	1	1	1	0	0	1
give-receive	1	0	0	0	1	0	0	1	0	1	1	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1	1	0	1	1
grow-improve	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1	0	1	1	0	1	0	1	1	0	0	1	1	0	1	0
hear-listen	1	1	1	1	0	1	1	0	0	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	0	1	1	1	1	1
hear-speak	1	1	1	1	0	1	0	0	0	1	0	0	1	1	1	1	0	0	0	1	1	1	1	0	0	1	1	1	0	1
join-meet	0	0	0	1	0	0	1	0	0	1	1	0	0	1	0	0	0	0	1	1	1	1	0	0	1	0	0	1	0	0
learn-read	1	0	0	1	1	1	0	0	0	1	1	0	0	0	0	1	0	1	0	0	0	1	1	1	0	0	0	1	0	0
listen-speak	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	0	1	1	1	0	1
read-write	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1
sell-spend	1	0	0	1	0	0	0	1	0	0	1	0	1	0	1	1	1	0	1	0	1	1	1	0	1	1	1	1	0	1
speak-tell	0	0	1	1	1	0	1	1	0	1	1	0	0	0	1	1	1	1	1	0	1	1	0	0	0	1	1	0	1	1

Note. 1 = presence; 0 = absence.

Appendix 6.1. Sorting task directions

The deck of cards in the envelope consists of 50 verbs. Please sort them into groups, by putting words with related meanings together. There might be a few words that don't seem to fit into any one group. You can leave these as single cards. It doesn't matter how many groups you make.

Please do it as quickly as you can. When you finish, please tell me "I'm done." Then I'll ask some simple questions.

You will get a small present when you finish.

Appendix 6.2a. Number of times words were associated with other words by NS (n = 30) (k = 50)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	*add*	30	2	2	3	2	4	5		1	2	1	1	5	3		4	7		7	3	2	7	22	8	2	2	1	5	2	1	1	1	2	8	1	2	3	3	2	2	2	1	3	1	2		8	1	2	1
2	*ask*	2	30			2	1		1	6	22	1	4	1	3		3			2	12	1		2	2		13	8	2			26	1				23	7	6	14		21	5						23	3	
3	*beat*	2		30		2	17	4	1	5		2	6	4	1	26	3	3	1	2	1	6	4		3	1	1	1	4	2	1		2	1	1	3			5	3	3		3			25	23	5	2	2	3
4	*begin*	3			30		1	4	14	1		2		5	1		3	4	14	2		2		2	2	15			4	1	10	1	6	15	4	7			3		3			29	17		1	2			6
5	*believe*	2	2	2		30	1			5	3		5		18	2	5			2	4	2	1	1	1		7	3	1	17		2			1		4	5	3	2	1	3	5			2	3	1	5	25	
6	*break*	4	1	17	1	1	30	6	1	3	1	5	4	8	2	16	4	4	1	3	3	7	7	1	4	3	1	3	10	1	3	1	5	3	4	6	2	2	7	6	7	4	4	1	2	17	13	6	2	1	6
7	*bring*	5		4	4		6	30	4			3	1	3	1	3	12	18	6	13	1	5	17	2	12	7		1	8	1	6		6	7	5	4	1	1	5	3	3	1	1	5	4	3	2	19	2		4
8	*come*		1	1	14		1	4	30	1		3	2	7		1	3	5	25	1		1	3		2	22		1	6		13		10	22	7	11	1	1	4		6	1	1	13	11	2		3	1		12
9	*cry*	1	6	5	1	5	3		1	30	7	6	4	4	5	7	1			2	3	3	2	1	1		5	3	2	7		7	3		3	4	7	1	4	9	5	10	3	1	1	6	10		7	4	3
10	*describe*	2	22			3	1			7	30	1	3	1	5		2			2	9	1		2	2		11	6	2	2		20	1				18	7	8	12		16	3				1		19	3	
11	*die*	1	1	2	2		5	3	3	6	1	30	5	8	3	3	4	3	2	2	1	3	4	2	3	2		3	3	2	4	1	12	2	6	4	2	2	4	2	7	3	2	2	8	3	4	2	1		4
12	*face*	1	4	6		5	4	1	2	4	3	5	30	3	3	6	6	1		2	2	7	3	2	2		3	8	3	5		3	4		2	2	4	6	10	3	3	4	9			4	7	3	5	4	2
13	*fall*	5	1	4	5		8	3	7	4	1	8	3	30	1	6	4	2	6	1	2	2	6	6	4	6		3	4		5	1	8	6	14	11	2	1	5	4	12	3	1	5	7	6	4	3			11
14	*feel*	3	3	1	1	18	2	1		5	5	3	3	1	30		6			3	8	2	2	2	2	1	7	7	2	16	2	3	2	1	1	1	6	11	1	4	2	4	8	1	2	1	2	1	4	17	1
15	*fight*			26		2	16	3	1	7		3	6	6		30	3	2	1		1	5	3	1	1	1	1	1	3	1	1		3	1	3	5			6	4	5	1	3			25	27	3	2	1	5
16	*find*	4	3	3	3	5	4	12	3	1	2	4	6	4	6	3	30	9	4	11	3	6	9	6	15	5	1	5	7	3	6	3	7	5	4	4	3	7	8	3	4	3	5	3	4	2	2	10	2	4	4
17	*get*	7		3	4		4	18	5			3	1	2		2	9	30	6	17		4	20	4	16	3			12	1	2		3	3	4	4			3	2	2		1	5	2	3	1	26	2		4
18	*go*			1	14		1	6	25			2		6		1	4	6	30	1		3	3		3	25			6		14	1	10	24	7	13			2	1	5			15	12	2		4	1		14
19	*have*	7	2	2	2	2	3	13	1	2	2	2	2	1	3		11	17	1	30	1	3	20	6	21			2	8	3	3	1	2				1	2	2			1	2	2	1	2	1	19		1	
20	*hear*	3	12	1		4	3	1		3	9	1	2	2	8	1	3			1	30		1	2	1	1	25	13		3	2	12	2	1	1	3	15	13	2	10	4	14	12		1	1			13	5	2
21	*help*	2	1	6	2	2	7	5	1	3	1	3	7	2	2	5	6	4	3	3		30	5		4	3	2	2	5	4	3	2	3	3	3	5		1	7	4	4	1	5	3	1	5	5	6	4	3	5
22	*hold*	7		4		1	7	17	3	2		4	3	6	2	3	9	20	3	20	1	5	30	3	20	2		1	10	2	4		5	2	6	6	1	1	4	3	7	2		1	2	6	2	23	1	1	6
23	*increase*	22	2		2	1	1	2		1	2	2	2	6	2	1	6	4		6	2		3	30	7	1		2	3		2	1	3	1	11	1	2	3	3	1	1	2	1	2	1		1	4			1
24	*keep*	8	2	3	2	1	4	12	2	1	2	3	2	4	2	1	15	16	3	21	1	4	20	7	30	3		2	9	2	5	2	4	3	3	3	2	2	5	1	4	2		2	3	4		17		1	3
25	*leave*	2		1	15		3	7	22			2		6	1	1	5	3	25		1	3	2	1	3	30		1	3		17	1	11	29	6	12	1	1	3	1	5	1		14	13	1		1			13
26	*listen*	2	13	1		7	1			5	11		3		7	1	1				25	2					30	12		6		13					16	12	3	8	1	14	11			1	2		17	9	
27	*look*	1	8	1		3	3	1	1	3	6	3	8	3	7	1	5			2	13	2	1	2	2	1	12	30		2	2	8	3	1	1	3	13	24	9	7	3	10	25		1	1			9	4	3
28	*make*	5	2	4	4	1	10	8	6	2	2	3	3	4	2	3	7	12	6	8		5	10	3	9	3			30	1	2		4	3	6	7	1	1	8	3	5	3	2	5	2	5	2	12	3	1	7
29	*mind*	2		2	1	17	1	1		7	2	2	5		16	1	3	1		3	3	4	2		2		6	2	1	30		1					3	4	2		1	2	4	1	1	2	4	2	3	16	
30	*remain*	1		1	10		3	6	13			4		5	2	1	6	2	14	3	2	3	4	2	5	17		2	2		30	1	21	17	5	6	1	2	2	1	12	1	2	9	18	1		1			7
31	*reply*	1	26		1	2	1			7	20	1	3	1	3		3		1	1	12	2		1	2	1	13	8		1	1	30	1	2			22	5	5	13		21	5	1	1				22	3	
32	*rest*	1	1	2	6		5	6	10	3	1	12	4	8	2	3	7	3	10	2	2	3	5	3	4	11		3	4		21	1	30	11	8	7	2	3	6	2	17	3	3	6	16	2	2	2	1		8
33	*return*	2		1	15		3	7	22			2		6	1	1	5	3	24		1	3	2	1	3	29		1	3		17	2	11	30	6	12	1	1	3	1	5	1		14	13	1		1			13
34	*rise*	8		1	4	1	4	5	7	3		6	2	14	1	3	4	4	7		1	3	6	11	3	6		1	6		5		8	6	30	11	1	1	4	3	11	3	1	5	5	3	2	3	1		11
35	*run*	1		3	7		6	4	11	4		4	2	11	1	5	4	4	13		3	5	6	1	3	12		3	7		6		7	12	11	30	1	2	5	7	18	4	3	8	6	6	3	4	1		28
36	*say*	2	23			4	2	1	1	7	18	2	4	2	6		3			1	15		1	2	2	1	16	13	1	3	1	22	2	1	1	1	30	10	6	16	1	27	8		1				25	5	1
37	*see*	3	7			5	2	1	1	1	7	2	6	1	11		7			2	13	1	1	3	2	1	12	24	1	4	2	5	3	1	1	2	10	30	8	5	3	7	23		1		1		7	6	2
38	*show*	3	6	5	3	3	7	5	4	4	8	4	10	5	1	6	8	3	2	2	2	7	4	3	5	3	3	9	8	2	2	5	6	3	4	5	6	8	30	5	5	6	9	3	1	5	5	5	7	2	5
39	*sing*	2	14	3		2	6	3		9	12	2	3	4	4	4	3	2	1		10	4	3	1	1	1	8	7	3		1	13	2	1	3	7	16	5	5	30	6	17	7	1	1	2	3	3	16	2	6
40	*sit*	2		3	3	1	7	3	6	5		7	3	12	2	5	4	2	5		4	4	7	1	4	5	1	3	5	1	12		17	5	11	18	1	3	5	6	30	4	4	3	11	6	3	3		1	17
41	*speak*	2	21			3	4	1	1	10	16	3	4	3	4	1	3			1	14	1	2	2	2	1	14	10	3	2	1	21	3	1	3	4	27	7	6	17	4	30	7		1	1	1		23	4	3
42	*stare*	1	5	3		5	4	1	1	3	3	2	9	1	8	3	5	1		2	12	5		1			11	25	2	4	2	5	3		1	3	8	23	9	7	4	7	30			2	2	1	9	5	3
43	*start*	3			29		1	5	13	1		2		5	1		3	5	15	2		3	1	2	2	14			5	1	9	1	6	14	5	8			3	1	3			30	16		1	3	1		7
44	*stop*	1			17		2	4	11	1		8		7	2		4	2	12	1	1	1	2	1	3	13		1	2	1	18	1	16	13	5	6	1	1	1	1	11	1		16	30	1	1	1			6
45	*strike*	2		25		2	17	3	2	6		3	4	6	1	25	2	3	2	2	1	5	6		4	1	1	1	5	2	1		2	1	3	6			5	2	6	1	2		1	30	22	5	1	2	6
46	*struggle*			23	1	3	13	2		10	1	4	7	4	2	27	2	1		1		5	2	1			2		2	4			2		2	3		1	5	3	3	1	2	1	1	22	30	2	2	2	3
47	*take*	8		5	2	1	6	19	3			2	3	3	1	3	10	26	4	19		6	23	4	17	1			12	2	1		2	1	3	4			5	3	3		1	3	1	5	2	30	2	1	4
48	*tell*	1	23	2		5	2	2	1	7	19	1	5		4	2	2	2	1		13	4	1				17	9	3	3		22	1		1	1	25	7	7	16		23	9	1		1	2	2	30	6	1
49	*think*	2	3	2		25	1			4	3		4		17	1	4			1	5	3	1		1		9	4	1	16		3					5	6	2	2	1	4	5			2	2	1	6	30	
50	*walk*	1		3	6		6	4	12	3		4	2	11	1	5	4	4	14		2	5	6	1	3	13		3	7		7		8	13	11	28	1	2	5	6	17	3	3	7	6	6	3	4	1		30

Note. Blank cells equal "0".

Appendix 6.2b. Number of times words were associated with other words by NNS (n = 30) (k = 50)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	*add*	30	4	1	2	2	4	4	1	2	5	2	3	7	1	1	3	1	1	3	1	4	5	19	4	4	2	2	6	2	5	4	4	4	9	2	2	1	5	2	4	2	2	2	2	2	2	1	3	2	2
2	*ask*	4	30	2		4	1	2	2	8	15	2	4	1	3	2	4	1	1	1	6	2	2	1	1	2	7	3	2	4	2	24	2	1	1	1	19	2	11	13	2	20	4			2	1	2	24	5	3
3	*beat*	1	2	30	1	1	19	4	3	5	3	10	8	4	1	26		3	3	1	1	4	1			3	2	1	3	1		4	3	3	2	4	2		3	3	3	3	2	2	2	23	19	3	3	1	5
4	*begin*	2		1	30		1	3	9	1	2	3	1	5		1	1	5	11	3	1	2	1	3	2	12	1	1	3		2		6	8	5	9	1	1	1	3	6			25	12	2	2	6			8
5	*believe*	2	4	1		30				7	3	1	5		21	1	9				4	3					3	3		27	5	4	1				3	4	2	2		3	4			1	4		2	27	
6	*break*	4	1	19	1		30	5		5	3	7	5	4		19	1	3	1	1	1	4	4	3	2	1	2	2	5		2	2	3	3	4	2	2	1	1	2	3	2	2	2	2	17	17	2	2		2
7	*bring*	4	2	4	3		5	30	5	1	2	1	2	2		3	1	11	4	11		2	11	2	5	5	1	1	5		4	4	3	9	5	6	2		3	3	3	3	2	4	4	3	3	18	3		7
8	*come*	1	2	3	9			5	30		1	2	1	8		3		4	26	2					1	20			3		3	1	8	20	7	20			1	2	6	1	1	10	13	3		6	1		22
9	*cry*	2	8	5	1	7	5	1		30	7	4	6	2	8	5	3	1	1		4	4	2			2	5	6		8	2	7	3	2	1	2	10	5	5	12	3	8	6	1	1	4	7		8	6	1
10	*describe*	5	15	3	2	3	3	2	1	7	30	2	5	3	3	2	3	3	1	3	5	2	3	3	2	3	6	5	3	4	3	16	4	1	2	2	17	4	13	12	3	16	5			4	2	2	19	5	2
11	*die*	2	2	10	3	1	7	1	2	4	2	30	6	5	2	11	1	1	3	1	1	3	4	1	1	5	2	2	2	1	2	2	8	4	4	3	2	1	2	3	7	1	2	5	5	9	9		2	1	2
12	*face*	3	4	8	1	5	5	2	1	6	5	6	30	2	6	8	3	2		1	2	6	2	1	1	1	3	8	2	6	3	7	2	1	2	1	4	6	7	3	4	5	10			10	10	2	4	4	2
13	*fall*	7	1	4	5		4	2	8	2	3	5	2	30	1	5	1	2	8	2		1	3	6	1	8			2		4		6	7	14	7			1	3	6			5	5	3	2	2	1		7
14	*feel*	1	3	1		21				8	3	2	6	1	30	1	5	2		1	9	1	1		1		8	8	1	20	6	3	1		1		5	9	3	4	1	5	7			3	3		3	22	
15	*fight*	1	2	26	1	1	19	3	3	5	2	11	8	5	1	30		3	3	1		5	1			3	1	1	2	1		3	3	3	2	4	1		2	2	3	2	2	2	2	22	22	3	2	1	5
16	*find*	3	4		1	9	1	1		3	3	1	3	1	5		30	4		4	1	6	2	2	4	1		5	4	7	3	3	2		1			6	6		1		6				2	1	2	8	
17	*get*	1	1	3	5		3	11	4	1	3	1	2	2	2	3	4	30	6	17	2	2	7	2	8	4	2	2	9		1	1	3	4	3	5	2	2	2	3	2	1	1	3	2	4	4	19	2	1	4
18	*go*	1	1	3	11		1	4	26	1	1	3		8		3		6	30	2	1				1	23	1	1	3		3		9	21	7	23	1	1		3	7			13	16	3	1	7			23
19	*have*	3	1	1	3		1	11	2		3	1	1	2	1	1	4	17	2	30		4	16	3	16	3			12		3	1	2	1	2	1			2	1	2					2	2	16	1		1
20	*hear*	1	6	1	1	4	1			4	5	1	2		9		1	2	1		30					1	29	13		5	1	6	1	1		1	10	16	4	7	1	10	11	1	1	1			6	6	
21	*help*	4	2	4	2	3	4	2		4	2	3	6	1	1	5	6	2		4		30	2	2	3	1	1	1	6	3		4	1	1		1	1		3	1	1	1	2			5	8	2	1	2	1
22	*hold*	5	2	1	1		4	11		2	3	4	2	3	1	1	2	7		16		2	30	4	17	3	1	1	5		10	2	6	3	4	1	1		2	2	6	1	2	2	3	1	2	9	2		1
23	*increase*	19	1		3		3	2			3	1	1	6			2	2		3		2	4	30	5	1			5		5		1	1	11				2		1			2	2	1	2	2	1		
24	*keep*	4	1		2		2	5	1		2	1	1	1	1		4	8	1	16		3	17	5	30	3			9		11	1	4	2	2				2		4			1	4	1	2	7	1		
25	*leave*	4	2	3	12		1	5	20	2	3	5	1	8		3	1	4	23	3	1	1	3	1	3	30	2	2	2		7	2	10	21	8	21	2	1	2	4	8	1	1	13	15	3	1	4	2		20
26	*listen*	2	7	2	1	3	2	1		5	6	2	3		8	1		2	1		29	1	1			2	30	14		4	1	7	2	2		2	11	15	5	8	2	11	12	1	1	2	1		7	5	1
27	*look*	2	3	1	1	3	2	1		6	5	2	8		8	1	5	2	1		13	1	1			2	14	30		4	1	4	2	2		2	7	27	11	4	2	7	27	1	1	2	1		4	5	1
28	*make*	6	2	3	3		5	5	3		3	2	2	2	1	2	4	9	3	12		6	5	5	9	2			30		2	2	2	2	2	1	1		3	1	2	1		2	2	3	3	9	2		2
29	*mind*	2	4	1		27				8	4	1	6		20	1	7				5	3					4	4		30	5	5	1				4	5	3	2		4	5			2	5		2	26	
30	*remain*	5	2		2	5	2	4	3	2	3	2	3	4	6		3	1	3	3	1		10	5	11	7	1	1	2	5	30	2	10	6	6	3	1	1	2	2	8	1	1	3	7			2	2	5	3
31	*reply*	4	24	4		4	2	4	1	7	16	2	7		3	3	3	1		1	6	4	2		1	2	7	4	2	5	2	30	1	2		1	19	3	11	13	1	20	6			4	3	1	21	5	2
32	*rest*	4	2	3	6	1	3	3	8	3	4	8	2	6	1	3	2	3	9	2	1	1	6	1	4	10	2	2	2	1	10	1	30	8	6	11	2	1	2	4	20	1	2	7	12	3	1	2	2	1	10
33	*return*	4	1	3	8		3	9	20	2	1	4	1	7		3		4	21	1	1	1	3	1	2	21	2	2	2		6	2	8	30	8	22	2	1	1	3	6	1	1	12	15	3	2	3	1		21
34	*rise*	9	1	2	5		4	5	7	1	2	4	2	14	1	2	1	3	7	2			4	11	2	8			2		6		6	8	30	6			1	2	5			7	8	2	1	2	1		6
35	*run*	2	1	4	9		2	6	20	2	2	3	1	7		4		5	23	1	1	1	1			21	2	2	1		3	1	11	22	6	30	2	1	1	4	9	1	1	11	15	4	2	4	1		28
36	*say*	2	19	2	1	3	2	2		10	17	2	4		5	1		2	1		10	1	1			2	11	7	1	4	1	19	2	2		2	30	7	11	19	2	27	6	1	1	2	1		23	4	1
37	*see*	1	2		1	4	1			5	4	1	6		9		6	2	1		16					1	15	27		5	1	3	1	1		1	7	30	10	4	1	7	24	1	1	1			3	6	
38	*show*	5	11	3	1	2	1	3	1	5	13	2	7	1	3	2	6	2		2	4	3	2	2	2	2	5	11	3	3	2	11	2	1	1	1	11	10	30	7	2	12	12			4	2	2	14	3	2
39	*sing*	2	13	3	3	2	2	3	2	12	12	3	3	3	4	2		3	3	1	7	1	2			4	8	4	1	2	2	13	4	3	2	4	19	4	7	30	4	17	3	2	1	3	1	2	16	2	3
40	*sit*	4	2	3	6		3	3	6	3	3	7	4	6	1	3	1	2	7	2	1	1	6	1	4	8	2	2	2		8	1	20	6	5	9	2	1	2	4	30	1	2	7	10	3	1	2	2		8
41	*speak*	2	20	3		3	2	3	1	8	16	1	5		5	2		1			10	1	1			1	11	7	1	4	1	20	1	1		1	27	7	12	17	1	30	8			3	1	1	24	4	2
42	*stare*	2	4	2		4	2	2	1	6	5	2	10		7	2	6	1			11	2	2			1	12	27		5	1	6	2	1		1	6	24	12	3	2	8	30	1		3	2	1	5	4	2
43	*start*	2		2	25		2	4	10	1		5		5		2		3	13		1		2	2	1	13	1	1	2		3		7	12	7	11	1	1		2	7		1	30	14	2	1	4			10
44	*stop*	2		2	12		2	4	13	1		5		5		2		2	16		1		3	2	4	15	1	1	2		7		12	15	8	15	1	1		1	10			14	30	2	1	2			14
45	*strike*	2	2	23	2	1	17	3	3	4	4	9	10	3	3	22		4	3	2	1	5	1	1	1	3	2	2	3	2		4	3	3	2	4	2	1	4	3	3	3	3	2	2	30	18	4	2	2	5
46	*struggle*	2	1	19	2	4	17	3		7	2	9	10	2	3	22	2	4	1	2		8	2	2	2	1	1	1	3	5		3	1	2	1	2	1		2	1	1	1	2	1	1	18	30	3	1	2	2
47	*take*	1	2	3	6		2	18	6		2		2	2		3	1	19	7	16		2	9	2	7	4			9		2	1	2	3	2	4			2	2	2	1	1	4	2	4	3	30	1		6
48	*tell*	3	24	3		2	2	3	1	8	19	2	4	1	3	2	2	2		1	6	1	2	1	1	2	7	4	2	2	2	21	2	1	1	1	23	3	14	16	2	24	5			2	1	1	30	3	2
49	*think*	2	5	1		27				6	5	1	4		22	1	8	1			6	2					5	5		26	5	5	1				4	6	3	2		4	4			2	2		3	30	
50	*walk*	2	3	5	8		2	7	22	1	2	2	2	7		5		4	23	1		1	1			20	1	1	2		3	2	10	21	6	28	1		2	3	8	2	2	10	14	5	2	6	2		30

Note. Blank cells equal "0".

Appendix 7.1. Comparison of dendrograms: Individual differences and goodness-of-fit against group data

	NS (n = 30)	NNS (n = 30)
Participant No.	dT_NS	dT_JP
1	107.61	70.54
2	65.84	46.96
3	52.87	117.95
4	114.85	89.17
5	161.81	77.24
6	49.70	118.92
7	65.48	134.39
8	103.20	110.73
9	56.43	78.72
10	90.00	140.75
11	100.15	116.64
12	205.88	96.41
13	65.95	96.60
14	159.09	115.41
15	71.86	69.91
16	158.75	141.56
17	68.89	41.75
18	211.42	146.78
19	82.61	174.26
20	92.01	61.29
21	113.73	52.49
22	44.41	122.64
23	189.00	116.54
24	109.90	233.84
25	107.30	111.25
26	73.82	58.93
27	146.19	143.53
28	68.36	109.19
29	91.41	114.84
30	81.20	77.18

Note: Dendrogram = Graphic representation (tree graph) of the results of a cluster analysis; dT_NS = the degree of goodness-of-fit of individual NS participants' dendrograms against the group's dendrogram as calculated by the square of the Minkowski distance between the distance matrices; dT_JP = the degree of goodness-of-fit of individual NNS (Japanese) participants' dendrograms against the group's dendrogram as calculated by the square of the Minkowski distance between the distance matrices.

Appendix 7.2. Distance of each individual participant's dendrogram for 1K adjective sorting task

Participant No.	NS (n = 30)	NNS (n = 30)
1	7	10
2	10	12
3	10	7
4	7	6
5	5	11
6	8	6
7	9	5
8	6	7
9	14	9
10	8	5
11	7	7
12	3	6
13	7	6
14	5	8
15	11	7
16	4	5
17	12	10
18	4	4
19	7	5
20	6	12
21	9	8
22	12	6
23	5	6
24	10	3
25	8	7
26	11	10
27	6	6
28	6	7
29	7	8
30	8	9

Note. The distance of an individual participant's dendrogram equals the number of cluster minus one (which excludes single, isolated words).

Appendix 7.3a. Number of times words were associated with other words by NS (n = 30) (k = 50)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	*afraid*	30	10		15		1		5	3	1	13		1	6		5	4	1	10	4		1	6		1				1	6		5	5	1	4	1		4		5		19	3	4	4	3	3	1		13
2	*alone*	10	30	1	4	1	2		1			12	1		1	1		1	3		3			2		4				1			1	7	2	1	1				19		8	3	2	1		1	3		3
3	*back*		1	30	1	3					15	1	7	18		28	2		5		2	12	1		4	12		4	5	12	3	16		2	2	1	7		5	3	1	4	1					1	5	6	1
4	*bad*	15	4	1	30	2	1		3	3	2	9		1	6	1	14	7	2	6	3	1	3	7	2		1	2		2	5	1	7	3	3	2	3		8		3	2	15	2	5	4	7	8	2	1	18
5	*big*		1	3	2	30		1		1	2	1	12	3	2	3	2		8		5	8	5	1	27	1	4	23	17	10	2	3	1	6	3		2		1	13	3	24	1	3				1	8	19	1
6	*black*	1	2		1		30	28		5		2											2			1	5				1			1				28			1										1
7	*blue*					1	28	30		6			1									1	2				5				2				1			28												1	
8	*certain*	5	1		3				30	6	2	2	1		6		4	7		6				5		1	1				5		5		3	10	2		11		1		4	4	9	25	13	3	2	1	11
9	*clear*	3			3	1	5	6	6	30	2	1	1	1	7		3	5	1	7	2	2	5	5	1		8	1	1	1	9	1	5	1	6	6	1	5	4	1	1	1	3	2	5	6	7	6	3	1	3
10	*close*	1		15	2	2			2	2	30		8	23		17	2	1	3	1		13		1	3	8	1	3	5	13	3	25	1	1	6	2	8		7	1		3	2		1	2	2		4	5	3
11	*dead*	13	12	1	9	1	2		2	1		30		1	2	1	2	1		1	5		2	3		4			1		4		1	15	1	4	1		2		8		9	4	3	3	2	4			9
12	*deep*		1	7		12		1	1	1	8		30	11		7			8		4	15	1	1	13	4	1	12	16	17	1	10	1	1	3		2		1	10	2	12					1		8	21	
13	*far*	1		18	1	3				1	23	1	11	30		18	1		4		3	17	1		5	10	1	5	9	16	2	26		2	5		6		5	4		5	1						4	8	1
14	*fine*	6	1		6	2			6	7		2			30		15	17		19	3	1	4	11	1	1	5	1	1		8		15	1	4	5	2		5	1	1	1	4	9	12	7	9	23	1	1	4
15	*forward*		1	28	1	3					17	1	7	18		30	2		5		2	12	1		4	12		4	5	13	3	18		2	2	2	8		5	3	1	4	1					1	5	6	1
16	*good*	5		2	14	2			4	3	2	2		1	15	2	30	20	1	16	2	1	3	15	2	1	3	2		1	9	1	18	3	3	1	1	1	11			2	6	6	13	5	12	19	3	1	9
17	*great*	4	1		7				7	5	1	1			17		20	30		19	2		2	14			4	1			8		18		4	2		1	6			1	6	12	22	7	11	19	3		4
18	*half*	1	3	5	2	8				1	3		8	4		5	1		30		2	5	1		9	2	3	10	8	8	2	4		2	3		3		1	10	8	10	2						22	8	1
19	*happy*	10			6				6	7	1	1			19		16	19		30			1	20			2				10		20	1	5	3	1	1	5				8	11	15	5	10	17	2		3
20	*hard*	4	3	2	3	5				2		5	4	3	3	2	2	2	2		30	4	9	1	5	2	9	4	6	4	2	2	2	9	3					8	3	4	5	6	2			2	3	5	2
21	*high*			12	1	8		1		2	13		15	17	1	12	1		5		4	30	3		10	7	2	10	15	25	3	15		3	3		4		2	8	1	10	1						5	14	1
22	*hot*	1		1	3	5	2	2		5		2	1	1	4	1	3	2	1	1	9	3	30	2	4	1	8	4	3	2	3		2	8		1	1	3		3	1	4	1	1	2		1	3	1	4	
23	*kind*	6	2		7	1			5	5	1	3	1		11		15	14		20	1		2	30	1	1	2	1			8		24	4	6	4	1	1	8		1	1	8	14	15	4	14	13	3	2	5
24	*large*			4	2	27				1	3		13	5	1	4	2		9		5	10	4	1	30	2	3	26	19	12	2	5	1	4	3		2		1	13	2	27	1	1					9	20	1
25	*last*	1	4	12		1	1		1		8	4	4	10	1	12	1		2		2	7	1	1	2	30		2	2	6	2	9		4	2	3	12		2	2	2	2	1			1		2	2	2	
26	*light*				1	4	5	5	1	8	1		1	1	5		3	4	3	2	9	2	8	2	3		30	5	2	2	2	1	3	3	2			6	1	5	1	5		5	3	1	2	2	3	2	
27	*little*			4	2	23				1	3		12	5	1	4	2	1	10		4	10	4	1	26	2	5	30	17	12	2	5	1	4	2		2		1	11	2	29	1	2					8	17	1
28	*long*			5		17				1	5	1	16	9	1	5			8		6	15	3		19	2	2	17	30	16	1	8		3	1		2			16	2	18							7	21	
29	*low*	1	1	12	2	10				1	13		17	16		13	1		8		4	25	2		12	6	2	12	16	30	2	15		3	2		5		3	10	2	12	2						6	17	1
30	*natural*	6		3	5	2	1	2	5	9	3	4	1	2	8	3	9	8	2	10	2	3	3	8	2	2	2	2	1	2	30	2	9	5	8	3	3	1	8	1		2	11	7	7	5	9	9	5	1	6
31	*near*			16	1	3				1	25		10	26		18	1		4		2	15			5	9	1	5	8	15	2	30		1	5	1	8		4	4		5	1						4	7	1
32	*nice*	5	1		7	1			5	5	1	1	1		15		18	18		20	2		2	24	1		3	1			9		30	2	6	2		1	8		1	1	8	12	17	4	12	16	3	2	6
33	*old*	5	7	2	3	6	1			1	1	15	1	2	1	2	3		2	1	9	3	8	4	4	4	3	4	3	3	5	1	2	30	1	1	4		1	2	7	4	6	6	1			3	2	3	2
34	*open*	1	2	2	3	3		1	3	6	6	1	3	5	4	2	3	4	3	5	3	3		6	3	2	2	2	1	2	8	5	6	1	30	4	3		4	4	1	2	4	6	4	4	5	4	6	4	3
35	*ready*	4	1	1	2				10	6	2	4			5	2	1	2		3			1	4		3					3	1	2	1	4	30	6		4				3	3	4	13	5	4			5
36	*recent*	1	1	7	3	2			2	1	8	1	2	6	2	8	1		3	1		4	1	1	2	12		2	2	5	3	8		4	3	6	30		1	1	1	2	2	1		2	1	2	2	1	1
37	*red*						28	28		5							1	1		1			3	1			6				1		1					30						1	1		1				
38	*right*	4		5	8	1			11	4	7	2	1	5	5	5	11	6	1	5		2		8	1	2	1	1		3	8	4	8	1	4	4	1		30			1	7	3	8	10	18	4	3	1	18
39	*round*			3		13				1	1		10	4	1	3			10		8	8	3		13	2	5	11	16	10	1	4		2	4		1			30	2	12		2					11	15	
40	*single*	5	19	1	3	3	1		1	1		8	2		1	1			8		3	1	1	1	2	2	1	2	2	2			1	7	1		1			2	30	2	4	4	3	1		2	6	2	2
41	*small*			4	2	24				1	3		12	5	1	4	2	1	10		4	10	4	1	27	2	5	29	18	12	2	5	1	4	2		2		1	12	2	30	1	2					8	18	1
42	*strange*	19	8	1	15	1			4	3	2	9		1	4	1	6	6	2	8	5	1	1	8	1	1		1		2	11	1	8	6	4	3	2		7		4	1	30	4	5	3	5	3	3		13
43	*strong*	3	3		2	3			4	2		4			9		6	12		11	6		1	14	1		5	2			7		12	6	6	3	1	1	3	2	4	2	4	30	15	6	7	10	2		3
44	*successful*	4	2		5				9	5	1	3			12		13	22		15	2		2	15			3				7		17	1	4	4		1	8		3		5	15	30	9	12	14	3		6
45	*sure*	4	1		4				25	6	2	3			7		5	7		5				4		1	1				5		4		4	13	2		10		1		3	6	9	30	11	7	2		10
46	*true*	3			7				13	7	2	2	1		9		12	11		10			1	14			2				9		12		5	5	1	1	18				5	7	12	11	30	8	2	1	15
47	*well*	3	1	1	8	1			3	6		4			23	1	19	19		17	2		3	13		2	2				9		16	3	4	4	2		4		2		3	10	14	7	8	30	2		3
48	*whole*	1	3	5	2	8			2	3	4		8	4	1	5	3	3	22	2	3	5	1	3	9	2	3	8	7	6	5	4	3	2	6		2		3	11	6	8	3	2	3	2	2	2	30	8	2
49	*wide*			6	1	19		1	1	1	5		21	8	1	6	1		8		5	14	4	2	20	2	2	17	21	17	1	7	2	3	4		1		1	15	2	18					1		8	30	
50	*wrong*	13	3	1	18	1	1		11	3	3	9		1	4	1	9	4	1	3	2	1		5	1			1		1	6	1	6	2	3	5	1		18		2	1	13	3	6	10	15	3	2		30

Note. Blank cells equal "0"

Appendix 7.3b. Number of times words were associated with other words by NNS (n = 30) (k = 50)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	*afraid*	30	6	1	18				8	4		14			6		5	5		12	7		1	6		1
2	*alone*	6	30		4	1			2	2		10	1		1		1	1	6	4	3		2	4	1	2
3	*back*	1		30	1					1	15	1	5	17		25			3			4				4
4	*bad*	18	4	1	30				6	5		12			12		14	7		7	8		2	7		1
5	*big*		1			30	1	1		4	3		15	3				5	13		12	18	8		28	
6	*black*					1	30	30		3		1											2			1
7	*blue*					1	30	30		3		1											2			1
8	*certain*	8	2		6				30	11		2			9	1	8	5		8	1		1	6		
9	*clear*	4	2	1	5	4	3	3	11	30	2	2	4	2	8	2	8	3	3	6	2	3	6	5	4	2
10	*close*			15		3				2	30		7	23		16			4		2	5	3	1	3	2
11	*dead*	14	10	1	12		1	1	2	2		30	1	1	2		2	1	1	5	6		2	4		4
12	*deep*		1	5		15				4	7	1	30	9		5		4	7		11	23	8		15	3
13	*far*			17		3				2	23	1	9	30		17			3		1	7	2		3	4
14	*fine*	6	1		12				9	8		2			30	3	24	17		19	2		4	17		
15	*forward*			25					1	2	16		5	17	3	30	4	3	3	4		4		3		4
16	*good*	5	1		14				8	8		2			24	4	30	20		21	2		3	20		
17	*great*	5	1		7	5			5	3		1	4		17	3	20	30	1	18	6	4	5	15	5	
18	*half*		6	3		13				3	4	1	7	3		3		1	30		5	10	3		15	1
19	*happy*	12	4		7				8	6		5			19	4	21	18		30	3		3	21		
20	*hard*	7	3		8	12			1	2	2	6	11	1	2		2	6	5	3	30	10	9	3	11	
21	*high*			4		18				3	5		23	7		4		4	10		10	30	6		19	1
22	*hot*	1	2		2	8	2	2	1	6	3	2	8	2	4		3	5	3	3	9	6	30	3	8	1
23	*kind*	6	4		7				6	5	1	4			17	3	20	15		21	3		3	30		
24	*large*		1			28				4	3		15	3				5	15		11	19	8		30	
25	*last*	1	2	4	1		1	1		2	2	4	3	4		4			1			1	1			30
26	*light*			1		18	4	4	1	5	2	1	12	1	1	1	1	1	11	1	9	14	8	1	18	1
27	*little*	2	5	1	2	21				2	3	2	11	2				3	11		8	12	7		22	1
28	*long*			1		23				3	3		17	5		1		5	12		10	22	7		23	2
29	*low*	1	1	5	1	14				2	4	1	16	5		4		1	9		8	23	5		15	1
30	*natural*	4	1		3	2	1	1	11	11	2	4	2		5	1	7	6	3	5	5	1	6	7	2	2
31	*near*			18		2				1	25		8	28		18			3		1	6	2		2	3
32	*nice*	5	1		11				7	7		1			23	4	27	23		21	3		4	21		
33	*old*	7	8	2	5	3	1	1	1	2		12	3	1	1	1	1	2	3	3	5	3	4	4	3	8
34	*open*	1	2	5	1	3			1	9	11	3	5	7	3	6	3	1	2	4	5	3	6	4	3	1
35	*ready*	4	5		3	1			7	4	1	5	2		7	2	7	7		8	2	1	4	7	1	2
36	*recent*	1	2	4						1	3	3	2	3		4			1	1				1		19
37	*red*					1	30	30		3		1											2			1
38	*right*	5	1	2	9				16	9	2	2		2	16	4	16	8	1	11	1		1	13		
39	*round*			2		16	2	2		4	5	2	8	5		2			11		8	11	5		14	
40	*single*	1	22		1	5	1	1	1	3	1	6	3		1		1	1	12	2	4	4	4	2	6	1
41	*small*		2	1		27	1	1		3	3		12	2				3	13		10	15	7		25	
42	*strange*	19	5		16				8	3		10			6		5	6		8	8		2	8		1
43	*strong*	4	5		5	5			1	2		5	4		6	3	8	13	4	9	11	5	4	12	5	
44	*successful*	7	3		7				8	5		4			16	4	17	12		17	3		2	17		
45	*sure*	8	2		6				28	12		2			11	2	10	6		10	1		1	8		
46	*true*	5		1	4				19	10	1		1	1	9	3	8	5	1	7	2		1	8		1
47	*well*	5	1		13				9	8		2			26	3	28	19		20	2		3	19		
48	*whole*		6	5		9				2	7		7	7		5			20		5	9	3		11	
49	*wide*			3		20				4	5		22	7	1	3	1	5	12	1	11	25	7	1	19	1
50	*wrong*	16	3	1	24				8	3		10			8		9	5	1	4	7		2	9		1

		26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	*afraid*		2		1	4		5	7	1	4	1		5		1		19	4	7	8	5	5			16
2	*alone*		5		1	1		1	8	2	5	2		1		22	2	5	5	3	2		1	6		3
3	*back*	1	1	1	5		18		2	5		4		2	2		1					1		5	3	1
4	*bad*		2		1	3		11	5	1	3			9		1		16	5	7	6	4	13			24
5	*big*	18	21	23	14	2	2		3	3	1		1		16	5	27		5					9	20	
6	*black*	4				1			1				30		2	1	1									
7	*blue*	4				1			1				30		2	1	1									
8	*certain*	1				11		7	1	1	7			16		1		8	1	8	28	19	9			8
9	*clear*	5	2	3	2	11	1	7	2	9	4	1	3	9	4	3	3	3	2	5	12	10	8	2	4	3
10	*close*	2	3	3	4	2	25			11	1	3		2	5	1	3					1		7	5	
11	*dead*	1	2		1	4		1	12	3	5	3	1	2	2	6		10	5	4	2		2			10
12	*deep*	12	11	17	16	2	8		3	5	2	2			8	3	12		4			1		7	22	
13	*far*	1	2	5	5		28		1	7		3		2	5		2					1		7	7	
14	*fine*	1				5		23	1	3	7			16		1		6	6	16	11	9	26		1	8
15	*forward*	1		1	4	1	18	4	1	6	2	4		4	2				3	4	2	3	3	5	3	
16	*good*	1				7		27	1	3	7			16		1		5	8	17	10	8	28		1	9
17	*great*	1	3	5	1	6		23	2	1	7			8		1	3	6	13	12	6	5	19		5	5
18	*half*	11	11	12	9	3	3		3	2		1		1	11	12	13		4			1		20	12	1
19	*happy*	1				5		21	3	4	8	1		11		2		8	9	17	10	7	20		1	4
20	*hard*	9	8	10	8	5	1	3	5	5	2			1	8	4	10	8	11	3	1	2	2	5	11	7
21	*high*	14	12	22	23	1	6		3	3	1				11	4	15		5					9	25	
22	*hot*	8	7	7	5	6	2	4	4	6	4		2	1	5	4	7	2	4	2	1	1	3	3	7	2
23	*kind*	1				7		21	4	4	7	1		13		2		8	12	17	8	8	19		1	9
24	*large*	18	22	23	15	2	2		3	3	1				14	6	25		5					11	19	
25	*last*	1	1	2	1	2	3		8	1	2	19	1			1		1				1			1	1
26	*light*	30	17	14	15	5	1	1	4	4			4	1	13	5	20		3	1	1	1	1	7	13	
27	*little*	17	30	16	14	2	2		6	3					10	5	24	1	3					7	12	1
28	*long*	14	16	30	16	1	4		4	3	1	1			17	4	20		5					8	23	
29	*low*	15	14	16	30	1	5		5	2		1			9	5	17	1	2					7	19	1
30	*natural*	5	2	1	1	30	1	8	5	7	4		1	9	3	2	2	8	6	7	12	12	7	2	1	7
31	*near*	1	2	4	5	1	30			8	1	3		2	4		2					1		6	6	
32	*nice*	1				8		30	2	3	6			13		1		7	10	15	9	8	25		1	8
33	*old*	4	6	4	5	5		2	30	1	3	9	1	2	3	5	4	7	7	2	1	2	1	2	3	6
34	*open*	4	3	3	2	7	8	3	1	30	4	1		2	5	2	3	1	2	5	2	3	3	3	4	1
35	*ready*			1		4	1	6	3	4	30	3		5		3		3	6	9	8	3	7		2	3
36	*recent*			1	1		3		9	1	3	30				1		1	1	1		1				
37	*red*	4				1			1				30		2	1	1									
38	*right*	1				9	2	13	2	2	5			30		1		5	5	11	18	20	17	2		13
39	*round*	13	10	17	9	3	4		3	5			2		30	5	15		2					11	15	
40	*single*	5	5	4	5	2		1	5	2	3	1	1	1	5	30	6	2	3	1	1		1	11	3	1
41	*small*	20	24	20	17	2	2		4	3			1		15	6	30		3					8	17	
42	*strange*		1		1	8		7	7	1	3	1		5		2		30	6	9	8	6	5			18
43	*strong*	3	3	5	2	6		10	7	2	6	1		5	2	3	3	6	30	9	2	4	7	1	5	6
44	*successful*	1				7		15	2	5	9	1		11		1		9	9	30	10	8	17		1	7
45	*sure*	1				12		9	1	2	8			18		1		8	2	10	30	21	11			8
46	*true*	1				12	1	8	2	3	3	1		20				6	4	8	21	30	9	1		10
47	*well*	1				7		25	1	3	7			17		1		5	7	17	11	9	30		1	9
48	*whole*	7	7	8	7	2	6		2	3				2	11	11	8		1			1		30	9	1
49	*wide*	13	12	23	19	1	6	1	3	4	2				15	3	17		5	1			1	9	30	
50	*wrong*		1		1	7		8	6	1	3			13		1		18	6	7	8	10	9	1		30

Note. Blank cells equal "0".

Appendix 8.1. Comparison of dendrograms: Individual differences and goodness-of-fit against group data

	NSs (n = 30)	NNSs (n = 30)
Participant No.	dT_NS	dT_JP
1	56.78	99.06
2	65.19	47.54
3	102.35	76.38
4	52.47	127.65
5	56.41	79.40
6	36.42	85.92
7	84.71	77.31
8	67.34	45.53
9	53.03	129.70
10	63.96	82.35
11	101.35	63.52
12	284.19	60.78
13	56.24	132.64
14	126.79	175.25
15	40.57	45.25
16	27.06	180.74
17	86.47	89.82
18	97.15	146.83
19	60.45	170.22
20	32.75	53.05
21	69.44	50.66
22	33.48	52.17
23	100.91	82.17
24	78.28	78.29
25	84.03	117.08
26	26.17	89.16
27	70.84	130.23
28	41.87	82.71
29	43.41	106.73
30	44.43	119.39

Note: Dendrogram = Graphic representation (i.e., tree graph) of the results of a cluster analysis; dT_NS = the degree of goodness-of-fit of individual NS participants' dendrograms against the group's dendrogram as calculated by the square of the Minkowski distance between the distance matrices; dT_JP = the degree of goodness-of-fit of individual NNS (Japanese) participants' dendrograms against the group's dendrogram as calculated by the square of the Minkowski distance between the distance matrices.

Appendix 8.2. Distance of each individual participant's dendrogram for 1K noun sorting task

Participant No.	NS (n = 30)	NNS (n = 30)
1	8	11
2	10	10
3	9	10
4	8	3
5	7	5
6	9	6
7	7	7
8	8	8
9	14	6
10	8	7
11	6	7
12	3	4
13	10	3
14	5	4
15	7	7
16	11	5
17	9	5
18	4	4
19	8	5
20	7	7
21	7	8
22	10	8
23	6	8
24	10	6
25	7	6
26	11	6
27	7	6
28	7	10
29	6	7
30	9	5

Note. The distance of an individual participant's dendrogram equals the number of cluster minus one (which excludes single, isolated words).

Appendix 8.3a: Number of times words were associated with other words by NS (n = 30) (k = 50)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	*air*	30	2		5		1			1	20			18				21	4	1			1		1	1	1	1		2		21			2		2	2	2	21		1		1	20	9	1	4	1	20	
2	*blood*	2	30	5				18	18	1		18			18	18			8		5		3		18		4			2			3	18		5			4				1			1	2				
3	*child*		5	30							1			1				1	7		29							1	1	1	1	1			1	29	2		2	1	1			1	1	1	1			1	
4	*color*	5			30	1		1	1		4	1		4	1		1	5	3	4			1			2	1			3		5	5		5		1	6	2	5	1	1		1	5	4	2	3	1	5	
5	*effect*				1	30	21	1	1	15		1			1	1	25		4	1			8	15	1	1		1	2	13	6		9	1	5		1	1	10		5	11				5	1		14		5
6	*effort*	1				21	30	1	1	15		1			1	1	19		4				7	15	1	1		1	4	18	8		9	1	4		1	1	9		5	11		2		6	1		13		4
7	*eye*		18		1	1	1	30	30	1		30			30	29	1						4	1	28		8	1	1				4	29	1			2	2		1	1	1			2	4		1		1
8	*face*		18		1	1	1	30	30	1		30			30	29	1						4	1	28		8	1	1				4	29	1			2	2		1	1	1			2	4		1		1
9	*fear*	1	1			15	15	1	1	30		1			1	1	16		7				10	14	1	1		1	2	11	4		12	1	3			1	12		4	15				4	2		20		3
10	*field*	20		1	4						30		3	28				24		6	1	4				1	4	1	1			24			4	1	1	1		25			3	3	23	10		5		23	
11	*foot*		18		1	1	1	30	30	1		30			30	29	1						4	1	28		8	1	1				4	29	1			2	2		1	1	1			2	4		1		1
12	*front*										3		30	3						10		29					13	1			2												28	1		3		18			
13	*ground*	18		1	4						28		3	30				22		6	1	4	1			2	4	1	1			22	1		4	1	1	1		23		1	3	3	21	10		5		21	
14	*hand*		18		1	1	1	30	30	1		30			30	29	1						4	1	28		8	1	1				4	29	1			2	2		1	1	1			2	4		1		1
15	*head*		18			1	1	29	29	1		29			29	30	1						4	1	28		8	1	1				3	30	1			1	2		1	1	1			2	3		1		1
16	*influence*				1	25	19	1	1	16		1			1	1	30		5				8	16	2	1		1	2	17	5		9	1	5		1	1	11		5	13				6	1		14		6
17	*island*	21		1	5						24			22				30	2	1	1					1	1			1		30			3	1	1	1		27				4	27	9		4		26	
18	*life*	4	8	7	3	4	4			7							5	2	30		7		10	4				1	1	10	2	2	8		2	7		1	18	2	1	9			2	3	1	1	9	1	1
19	*line*	1			4	1					6		10	6				1		30		11				3	8	1	2			1			7		1	2		2			10	2	1	4	1	8		1	1
20	*man*		5	29							1			1				1	7		30									1	1	1			1	30	1		2	1	1			1	1	1	1			1	
21	*middle*										4		29	4						11		30					13	1	1		2				1					1			27			3		18			
22	*mind*	1	3		1	8	7	4	4	10		4		1	4	4	8		10				30	7	5	3	2	1	2	7	4		18	4	1			1	19		5	21				6	2		17		6
23	*mistake*					15	15	1	1	14		1			1	1	16		4				7	30	2	2		1	3	11	6		6	1	4		1	2	7		6	10		1		6	2		9		5
24	*mouth*	1	18			1	1	28	28	1		28			28	28	2						5	2	30		7	2	1	1			3	28	1		1	2	2		1	2	1			2	4		1		2
25	*note*	1			2	1	1			1	1			2			1	1		3			3	2		30	4	5	4	2	2	1	4		7		15	12	2	1	9	3		1	1	2	10		1	1	11
26	*part*	1	4		1			8	8		4	8	13	4	8	8		1		8		13	2		7	4	30	1	1		2	1	1	8	1		3	1		1			15		1	1	2	10		1	
27	*party*	1		1		1	1	1	1	1	1	1	1	1	1	1	1		1	1		1	1	1	2	5	1	30	21		7		1	1	3		8	5	1		4	1	1	1		2	5	1	1		3
28	*play*			1		2	4	1	1	2	1	1		1	1	1	2		1	2		1	2	3	1	4	1	21	30		12		1	1	3		6	3	2	1	5	2				3	2		2		4
29	*power*	2	2	1	3	13	18			11							17	1	10		1		7	11	1	2				30	4	1	8		3	1	2	3	8	1	4	12		2	1	4	2	1	10	1	5
30	*rest*			1		6	8			4			2				5		2		1	2	4	6		2	2	7	12	4	30		2		1	1	2		4		5	5	2			3			5		1
31	*sea*	21		1	5						24			22				30	2	1	1					1	1			1		30			3	1	1	1		27				4	27	9		4		26	
32	*sense*		3		5	9	9	4	4	12		4		1	4	3	9		8				18	6	3	4	1	1	1	8	2		30	3	4			3	16		2	16				5	4		17		2
33	*shoulder*		18			1	1	29	29	1		29			29	30	1						4	1	28		8	1	1				3	30	1			1	2		1	1	1			2	3		1		1
34	*sign*	2		1	5	5	4	1	1	3	4	1		4	1	1	5	3	2	7	1	1	1	4	1	7	1	3	3	3	1	3	4	1	30	1	3	6	5	4	4	2		4	3	4	5	2	4	3	5
35	*son*		5	29							1			1				1	7		30									1	1	1			1	30	1		2	1	1			1	1	1	1			1	
36	*song*	2		2	1	1	1				1			1			1	1		1	1			1	1	15	3	8	6	2	2	1			3	1	30	19		1	14	1			1	1	20			1	14
37	*sound*	2			6	1	1	2	2	1	1	2		1	2	1	1	1	1	2			1	2	2	12	1	5	3	3		1	3	1	6		19	30	1	1	17	4			1	2	23		1	1	16
38	*spirit*	2	4	2	2	10	9	2	2	12		2			2	2	11		18		2		19	7	2	2		1	2	8	4		16	2	5	2		1	30		3	16				5	2		18		4
39	*star*	21		1	5						25			23				27	2	2	1	1				1	1			1		27			4	1	1	1		30				4	28	10		4		27	
40	*talk*			1	1	5	5	1	1	4		1			1	1	5		1		1		5	6	1	9		4	5	4	5		2	1	4	1	14	17	3		30	8				4	19		5		21
41	*thought*	1			1	11	11	1	1	15		1		1	1	1	13		9				21	10	2	3		1	2	12	5		16	1	2		1	4	16		8	30				7	5		22		7
42	*top*		1					1	1		3	1	28	3	1	1				10		27			1		15	1			2			1									30	1		3		19			
43	*train*	1		1	1		2				3		1	3				4		2	1			1		1		1		2		4			4	1				4			1	30	4	2		3		4	
44	*tree*	20		1	5						23			21				27	2	1	1					1	1			1		27			3	1	1	1		28				4	30	9		4		29	
45	*view*	9	1	1	4	5	6	2	2	4	10	2	3	10	2	2	6	9	3	4	1	3	6	6	2	2	1	2	3	4	3	9	5	2	4	1	1	2	5	10	4	7	3	2	9	30	1	5	5	9	2
46	*voice*	1	2	1	2	1	1	4	4	2		4			4	3	1		1	1	1		2	2	4	10	2	5	2	2			4	3	5	1	20	23	2		19	5				1	30		2		18
47	*west*	4			3						5		18	5				4	1	8		18					10	1		1		4			2					4			19	3	4	5		30		4	
48	*wonder*	1			1	14	13	1	1	20		1			1	1	14		9				17	9	1	1		1	2	10	5		17	1	4			1	18		5	22				5	2		30		5
49	*wood*	20		1	5						23			21				26	1	1	1					1	1			1		26			3	1	1	1		27				4	29	9		4		30	
50	*word*					5	4	1	1	3		1			1	1	6		1	1			6	5	2	11		3	4	5	1		2	1	5		14	16	4		21	7				2	18		5		30

Note. Blank cells equal "0".

Appendix 8.3b: Number of times words were associated with other words by NNS (n = 30) (k = 50)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	*air*	30		1	5	3	1				22			22			3	25	4	4	1					1		1	2	1	3	27			1	1	1	5		26	1		3	7	23	13	1	7	2	24	1
2	*blood*		30	8	1			24	24	1		25	3		25	24			7		8	1	3		24		5			2	1		4	25		8	2	2	4		1	1	1				11		1		
3	*child*	1	8	30				7	7	1	1	7		1	7	7		1	3	1	30	1	1		7		2	1				1	1	7		30	1	1	1	1	1	1		1	1		5		1	1	
4	*color*	5	1		30	6	2	2	2		7	1	1	7	1	1	6	6	6	6				2	2	7	2	5	4	2	3	6	2	1	8		5	9		5	5	2	3	3	7	6	4	4	2	7	7
5	*effect*	3			6	30	18			6	2			2			28	2	10	1			6	11		3	2	4	6	17	6	2	6		3		2	3	6	1	4	7	1	1	2	6	2		10	2	4
6	*effort*	1			2	18	30			14							19		9				13	14		3	1	3	7	18	6		13		3		3	2	13		5	15	1	1		5	2		16		6
7	*eye*		24	7	2			30	28	1		28	4		28	27			4		7	1	3		29		5						5	28		7	3	4	4		3	1	2			1	12		1		2
8	*face*		24	7	2			28	30	1		28	4		28	29			4		7	1	3		27		5						4	28		7	2	3	4		2	1	2				11		1		1
9	*fear*		1	1		6	14	1	1	30		1			1	1	7		7		1		24	7	1	3			2	11	4		21	1	1	1	1	1	24		2	23				6	2		23		5
10	*field*	22		1	7	2					30			30			2	25	3	9	1					1	1	2	3		3	24			2	1	1	5		22			3	8	24	12		6	2	25	
11	*foot*		25	7	1			28	28	1		30	3		30	28			4		7	1	3		27		5						4	30		7	2	2	4		1	1	2				10		1		
12	*front*		3		1			4	4			3	30		3	3			1	5		21	1		4		11			1	2		1	3	1			2	2		1		19	1			2	14			1
13	*ground*	22		1	7	2					30			30			2	25	3	9	1					1	1	2	3		3	24			2	1	1	5		22			3	8	24	12		6	2	25	
14	*hand*		25	7	1			28	28	1		30	3		30	28			4		7	1	3		27		5						4	30		7	2	2	4		1	1	2				10		1		
15	*head*		24	7	1			27	29	1		28	3		28	30			4		7	1	4		26		5						5	28		7	2	2	5		1	2	2				10		1		
16	*influence*	3			6	28	19			7	2			2			30	2	10	1			7	12		3	2	4	6	17	6	2	7		3		2	3	7	1	4	8	1	1	2	6	2		11	2	4
17	*island*	25		1	6	2					25			25			2	30	3	6	1							1	2		3	28			1	1		4		24			3	8	23	13		7	2	24	
18	*life*	4	7	3	6	10	9	4	4	7	3	4	1	3	4	4	10	3	30	1	3		7	5	4	5	4	7	8	12	12	3	7	4	4	3	5	3	8	2	7	7		2	3	4	6		8	3	6
19	*line*	4		1	6	1					9		5	9			1	6	1	30	1	5				5	4	2	1	1	1	6			7	1	2	4		5	1		6	5	5	2	1	9		6	4
20	*man*	1	8	30				7	7	1	1	7		1	7	7		1	3	1	30	1	1		7		2	1				1	1	7		30	1	1	1	1	1	1		1	1		5		1	1	
21	*middle*		1	1				1	1			1	21		1	1				5	1	30	1		1		12			1	2		1	1	1	1		1	1				20	1			1	14			
22	*mind*		3	1		6	13	3	3	24		3	1		3	4	7		7		1	1	30	6	3	2	1		2	10	4		23	3	1	1	1	3	28		2	24				4	4		21		6
23	*mistake*				2	11	14			7							12		5				6	30		4	1	4	10	9	10		7		2		2	1	6		6	8		1		4	1		6		6
24	*mouth*		24	7	2			29	27	1		27	4		27	26			4		7	1	3		30		5						5	27		7	4	5	4		4	1	1			1	13		1		2
25	*note*	1			7	3	3			3	1			1			3		5	5			2	4		30	1	7	4	2	2		6		14		11	9	1	1	10	4		3	1	2	6		4	1	17
26	*part*		5	2	2	2	1	5	5		1	5	11	1	5	5	2		4	4	2	12	1	1	5	1	30	2	4	3	4		2	5		2		1	1		3	1	11				3	7	1		1
27	*party*	1		1	5	4	3				2			2			4	1	7	2	1			4		7	2	30	17	4	7	1	1		5	1	12	7		2	10	1		4	2	2	5	1		2	7
28	*play*	2			4	6	7			2	3			3			6	2	8	1			2	10		4	4	17	30	7	15	2	3		2		9	5	2	3	14	5		3	2	6	3	2	4	2	5
29	*power*	1	2		2	17	18			11			1				17		12	1		1	10	9		2	3	4	7	30	8		6		2		1		10		5	9	2	2		3		1	11		6
30	*rest*	3	1		3	6	6			4	3		2	3			6	3	12	1		2	4	10		2	4	7	15	8	30	3	3		1		4		4	2	9	5	2	3	3	7		2	5	3	3
31	*sea*	27		1	6	2					24			24			2	28	3	6	1							1	2		3	30			1	1		4		26			3	8	24	13		8	2	25	
32	*sense*		4	1	2	6	13	5	4	21		4	1		4	5	7		7		1	1	23	7	5	6	2	1	3	6	3		30	4	3	1	4	6	23		6	21				7	7		19		9
33	*shoulder*		25	7	1			28	28	1		30	3		30	28			4		7	1	3		27		5						4	30		7	2	2	4		1	1	2				10		1		
34	*sign*	1			8	3	3			1	2		1	2			3	1	4	7		1	1	2		14		5	2	2	1	1	3		30		6	8	1	1	5	3	1	5	2	5	3	2	2	2	12
35	*son*	1	8	30				7	7	1	1	7		1	7	7		1	3	1	30	1	1		7		2	1				1	1	7		30	1	1	1	1	1	1		1	1		5		1	1	
36	*song*	1	2	1	5	2	3	3	2	1	1	2		1	2	2	2		5	2	1		1	2	4	11		12	9	1	4		4	2	6	1	30	19	1	1	18	2		3	2	2	18		1	1	12
37	*sound*	5	2	1	9	3	2	4	3	1	5	2	2	5	2	2	3	4	3	4	1	1	3	1	5	9	1	7	5			4	6	2	8	1	19	30	3	5	16	2	1	4	5	8	19	3	2	5	13
38	*spirit*		4	1		6	13	4	4	24		4	2		4	5	7		8		1	1	28	6	4	1	1		2	10	4		23	4	1	1	1	3	30		2	22				4	4		19		6
39	*star*	26		1	5	1					22			22			1	24	2	5	1					1		2	3		2	26			1	1	1	5		30	1		3	8	24	12	1	7	1	25	1
40	*talk*	1	1	1	5	4	5	3	2	2		1	1		1	1	4		7	1	1		2	6	4	10	3	10	14	5	9		6	1	5	1	18	16	2	1	30	6		1		4	14		4		17
41	*thought*		1	1	2	7	15	1	1	23		1			1	2	8		7		1		24	8	1	4	1	1	5	9	5		21	1	3	1	2	2	22		6	30				7	3		23		7
42	*top*	3	1		3	1	1	2	2		3	2	19	3	2	2	1	3		6		20			1		11			2	2	3		2	1			1		3			30	1	3	1		15		3	
43	*train*	7		1	3	1	1				8		1	8			1	8	2	5	1	1		1		3		4	3	2	3	8			5	1	3	4		8	1		1	30	10	6		4	1	9	2
44	*tree*	23		1	7	2					24			24			2	23	3	5	1					1		2	2		3	24			2	1	2	5		24			3	10	30	12		5	2	29	
45	*view*	13			6	6	5	1		6	12			12			6	13	4	2			4	4	1	2		2	6	3	7	13	7		5		2	8	4	12	4	7	1	6	12	30	3	5	7	12	3
46	*voice*	1	11	5	4	2	2	12	11	2		10	2		10	10	2		6	1	5	1	4	1	13	6	3	5	3				7	10	3	5	18	19	4	1	14	3				3	30		2		12
47	*west*	7			4						6		14	6				7		9		14					7	1	2	1	2	8			2			3		7			15	4	5	5		30	1	6	
48	*wonder*	2	1	1	2	10	16	1	1	23	2	1		2	1	1	11	2	8		1		21	6	1	4	1		4	11	5	2	19	1	2	1	1	2	19	1	4	23		1	2	7	2	1	30	2	5
49	*wood*	24		1	7	2					25			25			2	24	3	6	1					1		2	2		3	25			2	1	1	5		25			3	9	29	12		6	2	30	
50	*word*	1			7	4	6	2	1	5			1				4		6	4			6	6	2	17	1	7	5	6	3		9		12		12	13	6	1	17	7		2		3	12		5		30

Note. Blank cells equal "0".